Preprint : To appear in CVIU

Joint Radiometric Calibration and Feature Tracking for an Adaptive Stereo System

Seon Joo Kim [a] David Gallup [a] Jan-Michael Frahm [a]

Marc Pollefeys [a,b]

[a] *Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, N.C., U.S.A.*

[b] *Department of Computer Science, ETH Zurich, Zurich, Switzerland*

Abstract

To capture the full brightness range of natural scenes, cameras automatically adjust the exposure value which causes the brightness of scene points to change from frame to frame. Given such a video sequence, we introduce an adaptive stereo system that automatically generates texture-mapped 3D models. The key part of the system is a new method for tracking features and estimating the radiometric response function of the camera and the exposure difference between frames simultaneously. We model the global and nonlinear process that is responsible for the changes in image brightness rather than adapting to the changes locally and linearly which makes our tracking more robust to the change in brightness. The radiometric response function and the exposure difference between frames are also estimated in the process which enables our stereo process to deal with the varying brightness and generate radiometrically aligned textures.

Key words: radiometric calibration, stereo, KLT tracker, auto-exposure video

1 Introduction

Extracting and tracking features is a fundamental step in many computer vision systems since it provides means to relate one image to another spatially. One of the most commonly used feature trackers especially for processing videos is the KLT (Kanade-Lucas-Tomasi) tracker ([15,20]) due to its robustness and efficiency. However, there are cases that pose problems for the KLT tracker mainly when images of a high dynamic range scene are captured. In order to capture the full brightness range of natural scenes, where parts are in shadow and others are in bright sunlight for example, the camera has to adjust the exposure accordingly. In consequence, the appearance of the same scene point in the video sequence varies, breaking the basic assumption for the KLT tracker that the brightness of the scene points stay constant. Hence, we need methods to find the radiometric relationship between image features in addition to the spatial relationship.

We introduce an adaptive stereo system that automatically generates texture-mapped 3D models from videos captured with auto-exposure by developing a new tracking method which models the changes in image brightness between images globally and nonlinearly rather than treating the variation locally and linearly by comparing local regions independently [1]. The brightness change can be explained by the radiometric response function which defines the mapping from the image irradiance to the image brightness. We first introduce a method for tracking features and estimating the exposure changes between frames when the camera response function is known. In many cases the radiometric response function is not known in prior, so we present a method for

[1] Part of this work was introduced in [10]

joint feature tracking and radiometric calibration by formulating the estimation of the response function within a linear feature tracking scheme that can deal with varying intensity values of features due to exposure changes. Our novel frame work performs an integrated radiometric calibration in contrast to previous radiometric calibration techniques which require the correspondences as an input to the system which leads to a chicken-and-egg problem as precise tracking requires accurate radiometric calibration. By combining both into an integrated approach we solve this chicken-and-egg problem. With our joint estimation, we can advance the quality and robustness of the known structure from motion techniques by incorporating the information for 3D camera tracking, the depth from stereo and providing radiometrically aligned images for texture-mapping.

The remainder of the paper is organized as follows. In the next section, we will review the basics of the KLT tracker and the radiometric calibration process as well as the related works. In Section 3, we will first introduce a method for tracking features when the response function is known and then explain the method for simultaneous tracking and the response function estimation. We introduce our stereo method in Section 4 and evaluate our method with experiments in Section 5. We conclude with discussion about our algorithm and future works in Section 6.

2 Basics and Previous Works

2.1 Kanade-Lucas-Tomasi (KLT) Tracker

We first review the KLT tracker ([15,20]). The algorithm is based on the assumptions that the motion of the camera is small and the appearance of

features stays constant between consecutive frames in the video sequence. The brightness constancy assumption is stated as follows :

$$J(\mathbf{x} + \frac{\mathbf{dx}}{2}) - I(\mathbf{x} - \frac{\mathbf{dx}}{2}) = 0 \tag{1}$$

J and I are images at time $t+1$ and t respectively, $\mathbf{x} = [x, y]^T$ is the feature location, and $\mathbf{dx} = [dx, dy]^T$ is the displacement vector. After linearizing the equation and minimizing the error over a patch P, the displacement for each feature is computed as follows [2]:

$$\sum_{\mathbf{x} \in P} \begin{bmatrix} {s_x}^2 & s_x s_y \\ s_x s_y & {s_y}^2 \end{bmatrix} \begin{bmatrix} dx \\ dy \end{bmatrix} = 2 \sum_{\mathbf{x} \in P} \begin{bmatrix} (I(\mathbf{x}) - J(\mathbf{x}))s_x \\ (I(\mathbf{x}) - J(\mathbf{x}))s_y \end{bmatrix} \tag{2}$$

$$s_x = J_x + I_x, \quad s_y = J_y + I_y \tag{3}$$

where $J_x = \frac{\partial J(\mathbf{x})}{\partial x}$, $J_y = \frac{\partial J(\mathbf{x})}{\partial y}$, $I_x = \frac{\partial I(\mathbf{x})}{\partial x}$, and $I_y = \frac{\partial I(\mathbf{x})}{\partial y}$. Notice that the summation is over the patch P surrounding the feature due to the assumption of locally identical motion of features.

The dynamic range of cameras is usually too small to accommodate the large dynamic range of natural scenes. Accordingly, the exposure of the camera is adjusted causing the appearance of the features to change. In the implementation by Birchfield ([2]), a simple method is used to account for the gain change between images. For each feature patch P in the first image, an individual gain is computed using the current estimate of the location of the patch P' in the second image. The gain ratio is computed by the ratio of mean intensity values of the two patches. The estimated ratio is used to normalize the intensity of the neighborhoods of the point in the second image to proceed with the tracking

process. In [1,8] illumination invariance is also achieved by solving for a gain and bias factor in each individually tracked patch. In all of these approaches, the change in intensity is treated locally for each individual feature. Also, the intensity change which is a nonlinear process is linearly approximated.

2.2 Radiometric Calibration

The radiometric response function defines the relationship between the image irradiance E and the image brightness I at a point $\mathbf{x}$ in regards to the exposure k as follows.

$$I_{\mathbf{x}} = f(kE_{\mathbf{x}}) \tag{4}$$

Assuming equal irradiance for corresponding points between images [2], we have the following equation which shows how corresponding points $\mathbf{x_1}, \mathbf{x_2}$ in two images I, J are radiometrically related.

$$g(J(\mathbf{x_2})) - g(I(\mathbf{x_1})) = K \tag{5}$$

The response function g is the log-inverse of the radiometric response function [3] ($\log f^{-1}$) and the K is the logarithm of the exposure ratio between the two images. For simplicity, we will just consider g as the response function and K as the exposure difference between two images.

While there are several methods that do not require correspondences such as estimating the response function from the relation between histograms ([5]) or from a single image using color distribution of local edge regions ([13,14]), ma-

[2] We assume that the effect of vignetting is the same for the corresponding points since the displacement between consecutive frames is small in videos.
[3] See [12] for proof of invertability

jority of work relies on correspondences to compute the response function. In most cases, the correspondence problem is solved by taking images of a static scene with a static camera with different exposures ([4,17,18,6]). Then the response function is estimated by solving Eq. (5) in least squares sense using different models for the response function such as gamma curve ([17]), polynomial ([18]), non-parametric ([4]), and PCA model ([6]). While the restriction on the camera motion was relieved in [12] by dealing with large amount of outliers robustly, it relies on the computation of dense correspondence maps and assumes that the normalized-cross correlation used as the similarity measure can match features reliably over brightness changes.

2.3 Joint Domain and Range Image Registration

Similar to our work is the joint domain and range registration of images. In [16], Mann introduced a method for jointly computing the projective coordinate transform (domain) and the brightness change (range) between a pair of images of a static scene taken with a purely rotating camera. The brightness transform which they call the comparametric function is approximated by using a gamma function as the model for the response function. The joint registration process is linearized by the Taylor expansion and the least squares solution is acquired. In a similar work, Candocia proposed a method for the joint registration by using a piecewise linear model for the comparametric function ([3]).

Notice that our method is different from the methods [16,3] in that we are interested in tracking of features which are allowed to move unconstrained rather than a global projective transform between images. This involves esti-

mation of a significantly more parameters and our algorithm is able to deal with it efficiently. In addition, we do not restrict the movement of the camera and we can also deal with moving objects in the scene. We are also different in that we compute the actual response function of the camera and the exposures rather than just finding out the brightness transform between images.

2.4 Brightness Invariant Stereo

Several stereo methods have employed matching metrics which achieve invariance to brightness changes. A comparison of these techniques is presented in [7]. Normalized-cross correlation is effective for dealing with locally linear changes, while mutual information is invariant to arbitrary one-to-one mappings. However, mutual information has been only been successfully implemented as a global mapping between images. The rank transform has been shown to be robust even to local illumination changes. In general, more invariance can lead to ambiguity in some cases, and overfitting is possible. Furthermore, all these methods require known camera poses, or at least rectified images. In our system, we recover the camera poses from feature tracks. Using our method, the radiometric calibration is recovered jointly with the feature tracks, and therefore brightness invariance in stereo is unnecessary. We employ the more constrained absolute difference metric, after first normalizing the image intensities for radiometric differences. This extends our previous work [11] which accounted only for changes in exposure (gain) assuming a linear camera response.

3 Joint Feature Tracking and Radiometric Calibration

We now introduce our method for brightness invariant feature tracking and radiometric calibration. Given a video sequence with varying exposure, we

estimate the radiometric response function of the camera, the exposure difference between frames, and the feature tracks from frame to frame. Our feature tracking in contrast to previous approaches models the global and nonlinear process that is responsible for changes in image brightness rather than adapting to the changes locally and linearly. Our radiometric calibration is different from previous calibration works because the correspondences are output of our system rather than being an input to the system. Our method is an online process not a batch process which allows subsequent algorithms such as stereo matching to compensate for brightness changes.

We will first start with explaining the method for tracking features when the response function is known and then we will proceed to the method for the joint feature tracking and radiometric calibration.

3.1 Tracking Features with Known Response

We first explain the method for tracking features and estimating the exposure difference K between two images when the response function of the camera g is known. For a feature $\mathbf{x}$ with the displacement $\mathbf{dx}$, Eq. (5) becomes

$$g(J(\mathbf{x}+\frac{\mathbf{dx}}{2}))-g(I(\mathbf{x}-\frac{\mathbf{dx}}{2}))=K. \tag{6}$$

We apply the Taylor expansion to the images (Eq. (7)) and then to the response function (Eq. (8)) to linearize the equation above.

$$g(J(\mathbf{x})+\nabla J(\mathbf{x})^T\frac{\mathbf{dx}}{2})-g(I(\mathbf{x})-\nabla I(\mathbf{x})^T\frac{\mathbf{dx}}{2})=K \tag{7}$$

Let $J(\mathbf{x}) = J$, $I(\mathbf{x}) = I$, and g' be the derivative of the response function g,

$$g(J) + g'(J)\nabla J^T \frac{\mathbf{dx}}{2} - \left[g(I) - g'(I)\nabla I^T \frac{\mathbf{dx}}{2} \right] - K = 0 \tag{8}$$

Assuming equal displacement for all pixels of a patch around each feature P_i, the displacements for each feature $[dx_i, dy_i]^T$ and the exposure difference K are estimated by minimizing the following error function :

$$E(dx_i, dy_i, K) = \sum_{\mathbf{x} \in P_i} (\beta + a\frac{dx_i}{2} + b\frac{dy_i}{2} - K)^2 \tag{9}$$

with

$$a = g'(J(\mathbf{x}))J_x + g'(I(\mathbf{x}))I_x \tag{10}$$

$$b = g'(J(\mathbf{x}))J_y + g'(I(\mathbf{x}))I_y \tag{11}$$

$$\beta = g(J(\mathbf{x})) - g(I(\mathbf{x})) \tag{12}$$

The error function is minimized when all partial derivatives towards the unknowns are zero. Accordingly, the following equation needs to be solved for each feature.

$$\underbrace{\begin{bmatrix} \mathbf{U_i} & \mathbf{w_i} \\ \mathbf{w_i}^T & \lambda_i \end{bmatrix}}_{\mathbf{A_i}} \mathbf{z_i} = \begin{bmatrix} \mathbf{v_i} \\ m_i \end{bmatrix} \tag{13}$$

where,

$$\mathbf{U_i} = \begin{bmatrix} \frac{1}{2}\sum_{P_i} a^2 & \frac{1}{2}\sum_{P_i} ab \\ \frac{1}{2}\sum_{P_i} ab & \frac{1}{2}\sum_{P_i} b^2 \end{bmatrix} \tag{14}$$

$$\mathbf{w_i} = \begin{bmatrix} -\sum_{P_i} a \\ \\ -\sum_{P_i} b \end{bmatrix}, \quad \lambda_i = \sum_{P_i} 2 \tag{15}$$

$$\mathbf{v_i} = \begin{bmatrix} -\sum_{P_i} \beta a \\ \\ -\sum_{P_i} \beta b \end{bmatrix}, \quad m_i = 2 \sum_{P_i} \beta \tag{16}$$

$$\mathbf{z_i} = [dx_i, dy_i, K]^T \tag{17}$$

Note that the exposure difference K is global for all features and we can estimate the unknown displacements for all features (dx_i, dy_i, $i = 1$ to n) and the exposure K simultaneously by minimizing the following error.

$$E(dx_1, dy_1, ..., dx_n, dy_n, K) = \sum_{i=1}^{n} E(dx_i, dy_i, K) \tag{18}$$

Accordingly the unknowns are found by solving the following linear equation.

$$\mathbf{Az} = \begin{bmatrix} \mathbf{U} & \mathbf{w} \\ \\ \mathbf{w}^T & \lambda \end{bmatrix} \mathbf{z} = \begin{bmatrix} \mathbf{v} \\ \\ m \end{bmatrix} \tag{19}$$

with

$$\mathbf{U} = \begin{bmatrix} \mathbf{U_1} & 0 & \dots & 0 \\ 0 & \mathbf{U_2} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & & \dots & \mathbf{U_n} \end{bmatrix}, \quad \mathbf{w} = [\mathbf{w_1}, \dots, \mathbf{w_n}]^T \tag{20}$$

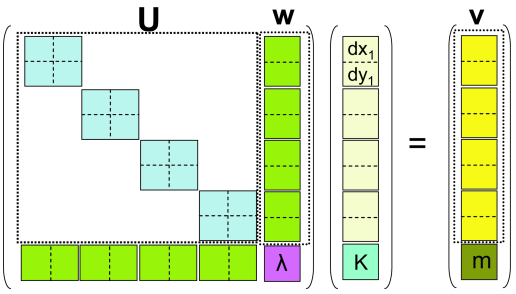

Fig. 1. Solving for the displacements and the exposure : Illustration of Eq. (19)

$$\lambda = \sum_{i=1}^{n} \lambda_i, \quad m = \sum_{i=1}^{n} m_i, \quad \mathbf{v} = [v_1, \ldots, v_n]^T \tag{21}$$

$$\mathbf{z} = [dx_1, dy_1, \ldots, dx_n, dy_n, K]^T \tag{22}$$

Fig. 1 shows the structure of Eq. (19). The matrix $\mathbf{A}$ is a sparse matrix and we can take advantage of its structure to find a computationally efficient solutions. Both sides of the Eq. (19) are multiplied on the left by $\begin{bmatrix} \mathbf{I} & 0 \\ -\mathbf{w}^T\mathbf{U}^{-1} & 1 \end{bmatrix}$ resulting in

$$\begin{bmatrix} \mathbf{U} & \mathbf{w} \\ \mathbf{0} & -\mathbf{w}^T\mathbf{U}^{-1}\mathbf{w} + \lambda \end{bmatrix} \mathbf{z} = \begin{bmatrix} \mathbf{v} \\ -\mathbf{w}^T\mathbf{U}^{-1}\mathbf{v} + m \end{bmatrix} \tag{23}$$

where $(-\mathbf{w}^T\mathbf{U}^{-1}\mathbf{w} + \lambda)$ is the Schur complement [21] of the matrix $\mathbf{U}$. Since the inverse of $\mathbf{U}$ can be computed efficiently as it is a 2×2 block diagonal matrix (this inversion corresponds to the amount of work necessary for the standard KLT) and its Schur complement is a scalar, Eq. (23) can be solved very efficiently. The exposure difference K is given by

$$(-\mathbf{w}^T\mathbf{U}^{-1}\mathbf{w} + \lambda)K = -\mathbf{w}^T\mathbf{U}^{-1}\mathbf{v} + m \tag{24}$$

Once K is found, we can solve for the displacements. For each patch i, dx_i and

dy_i are computed by back substituting K as in Eq. (25). Hence the proposed estimation adds one additional equation Eq. (24) to solve to the standard KLT tracking equations.

$$\mathbf{U_i} \begin{bmatrix} dx_i \\ \\ dy_i \end{bmatrix} = \mathbf{v_i} - K\mathbf{w_i} \tag{25}$$

3.2 Joint Tracking and Radiometric Calibration

We now discuss the case of unknown response function. Given a video sequence, we automatically compute the radiometric response function g, the exposure difference between frames K, and the feature tracks.

We use the Empirical Model of Response (EMoR) introduced by Grossberg and Nayar in [6]. They combined the theoretical space of the response function and the database of real world camera response functions to create the EMoR which is a Mth order approximation :

$$g(I) = \mathbf{g}_0(I) + \sum_{k=1}^{M} c_k \mathbf{h}_k(I) \tag{26}$$

where g_0 is the mean function and c_k's are the coefficients for the basis functions $\mathbf{h}_k$'s. In this paper, we used a third order approximation ($M = 3$) since the first three basis functions explain more than 99.6% of the energy ([6]). The derivative of the response function is similarly a linear combination of the derivatives of the basis functions.

$$g'(I) = \mathbf{g}'_0(I) + \sum_{k=1}^{M} c_k \mathbf{h}'_k(I) \tag{27}$$

Substituting g and g' in Eq. (8) with Eq. (26) and Eq. (27), we get the following equation.

$$d + a \cdot dx + b \cdot dy + \sum_{k=1}^{M} c_k r_k + \sum_{k=1}^{M} \alpha_k p_k + \sum_{k=1}^{M} \beta_k q_k - K = 0 \tag{28}$$

The known variables for Eq. (28) are :

$$a = \frac{g'_0(J)J_x + g'_0(I)I_x}{2}, \quad b = \frac{g'_0(J)J_y + g'_0(I)I_y}{2} \tag{29}$$

$$r_k = h_k(J) - h_k(I), \quad p_k = \frac{h'_k(J)J_x + h'_k(I)I_x}{2} \tag{30}$$

$$q_k = \frac{h'_k(J)J_y + h'_k(I)I_y}{2}, \quad d = g_0(J) - g_0(I) \tag{31}$$

The unknowns are the displacements dx and dy, the coefficients for the response function c_k ($k = 1$ to M), the exposure difference K, and variables introduced for linearization $\alpha_k = c_k dx$ and $\beta_k = c_k dy$.

Again, we assume constant displacement for all pixels in a patch around each feature and minimize the following error function to solve for the unknowns.

$$E(dx_i, dy_i, c_1, \ldots, c_M, \alpha_{i1}, \ldots, \alpha_{iM}, \beta_{i1}, \ldots, \beta_{iM}, K) =$$

$$\sum_{P_i} (d + a dx_i + b dy_i + \sum_{k=1}^{M} c_k r_k + \sum_{k=1}^{M} \alpha_{ik} p_k + \sum_{k=1}^{M} \beta_{ik} q_k - K)^2 \tag{32}$$

Setting all partial derivatives towards the unknowns to zero, we get following equation for each feature.

$$\underbrace{\begin{bmatrix} \mathbf{U_i} & \mathbf{W_i} \\ \mathbf{W_i}^T & \mathbf{\Lambda_i} \end{bmatrix}}_{\mathbf{A_i}} \mathbf{z_i} = \begin{bmatrix} \mathbf{v_i} \\ \mathbf{m_i} \end{bmatrix} \tag{33}$$

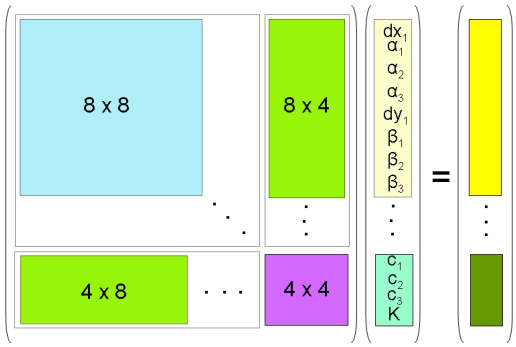

Fig. 2. Solving for the radiometric response function (3 basis functions), exposures, and the feature displacements : Illustration of Eq. (34).

Now we can solve for all feature tracks and the global parameters for the response function and the exposure difference similar to the case of known response function.

$$\mathbf{Az} = \begin{bmatrix} \mathbf{U} & \mathbf{W} \\ \mathbf{W}^T & \Lambda \end{bmatrix} \mathbf{z} = \begin{bmatrix} \mathbf{v} \\ \mathbf{m} \end{bmatrix} \tag{34}$$

with

$$\mathbf{U} = \begin{bmatrix} \mathbf{U_1} & 0 & \dots & 0 \\ 0 & \mathbf{U_2} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & & \dots & \mathbf{U_n} \end{bmatrix}, \quad \mathbf{W} = [\mathbf{W_1}, \dots \mathbf{W_n}]^T \tag{35}$$

$$\mathbf{\Lambda} = \sum_{i=1}^{n} \mathbf{\Lambda_i}, \quad \mathbf{v} = [\mathbf{v_1}, \dots, \mathbf{v_n}]^T, \mathbf{m} = \sum_{i=1}^{n} \mathbf{m_i} \tag{36}$$

$$\mathbf{z} = [\varphi_1, \dots, \varphi_n, c_1, \dots, c_M, K]^T \tag{37}$$

where

$$\varphi_i = [dx_i, \alpha_{i1}, \dots, \alpha_{iM}, dy_i, \beta_{i1}, \dots, \beta_{iM}]^T \tag{38}$$

Notice that Eq. (34) has the same structure as Eq. (19) (Fig. 2) except that the size of each sub-matrices are bigger. $\mathbf{U_i}$'s are $(2M+2) \times (2M+2)$, $\mathbf{W_i}$'s are $(2M+2) \times (M+1)$, and $\mathbf{\Lambda_i}$'s are $(M+1) \times (M+1)$. Multiplying both

sides on the left by $\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{W}^T\mathbf{U}^{-1} & \mathbf{I} \end{bmatrix}$ results in

$$\begin{bmatrix} \mathbf{U} & \mathbf{W} \\ \mathbf{0} & -\mathbf{W}^T\mathbf{U}^{-1}\mathbf{W} + \Lambda \end{bmatrix} \mathbf{z} = \begin{bmatrix} \mathbf{v} \\ -\mathbf{W}^T\mathbf{U}^{-1}\mathbf{v} + \mathbf{m} \end{bmatrix} \tag{39}$$

The coefficients of the response function and the exposure can be solved by

$$(-\mathbf{W}^T\mathbf{U}^{-1}\mathbf{W} + \Lambda)\upsilon = -\mathbf{W}^T\mathbf{U}^{-1}\mathbf{v} + \mathbf{m} \tag{40}$$

where

$$\upsilon = [c_1, \ldots, c_M, K] \tag{41}$$

The solution to Eq. (40) will suffer from the exponential ambiguity (or γ ambiguity) which means that if a response function g and an exposure K is the solution to the problem so are γg and γK [5]. Simply put, there are many response functions and exposures that satisfy the equation that are of different scales. As stated in [5], we have to make assumptions on either the response function or the exposure to fix the scale. To deal with this ambiguity problem, we chose to set the value of the response function at the image value at 128 to a value τ. This is done by adding the following equation to Eq. (40).

$$\omega \sum_{k=1}^{M} c_k \mathbf{h_k}(128) = \omega(\tau - \mathbf{g_0}(128)) \tag{42}$$

The value ω in the equation controls the strength of the constraint.

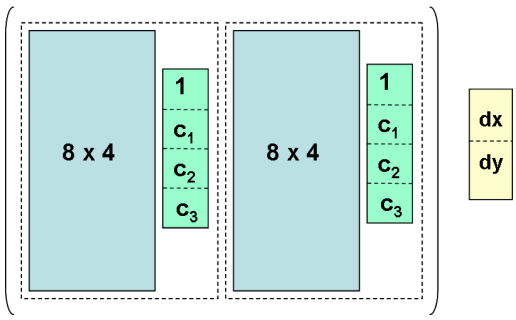

Fig. 3. Factorization for estimating the tracks.

The displacement for each feature can then solved by back substituting the solution υ to Eq. (33).

$$\mathbf{U_i}\varphi_i = \mathbf{v_i} - \mathbf{W_i}\upsilon \tag{43}$$

Notice that α_{ik}'s and β_{ik}'s in φ_i are the products of the displacement and the response function coefficients : $\alpha_{ik} = c_k dx_i$ and $\beta_{ik} = c_k dy_i$. Since we have already estimated the coefficients c_k's, Eq. (43) can be factorized in a simpler form as follows Fig. 3.

$$\begin{bmatrix} dx_i & \alpha_{i1} & \dots & \alpha_{iM} \\ dy_i & \beta_{i1} & \dots & \beta_{iM} \end{bmatrix} = \begin{bmatrix} dx_i \\ dy_i \end{bmatrix} \underbrace{\begin{bmatrix} 1 & c_1 & \dots & c_M \end{bmatrix}}_{\underline{\mathbf{c}}} \tag{44}$$

$$\mathbf{Y}_i \begin{bmatrix} dx_i \\ dy_i \end{bmatrix} = \mathbf{v_i} - \mathbf{W_i}\upsilon \tag{45}$$

$$\begin{aligned} \mathbf{Y}_i(\cdot, 1) &= \mathbf{U}_i(\cdot, 1 : M+1)\underline{\mathbf{c}}^T \\ \mathbf{Y}_i(\cdot, 2) &= \mathbf{U}_i(\cdot, M+2 : 2M+2)\underline{\mathbf{c}}^T \end{aligned} \tag{46}$$

3.3 Updating the Response Function Estimate

In Section 3.2, we introduced the method for computing the response function, the exposure difference, and the feature tracks at the same time given an image pair from a video sequence. We now explain how we can integrate the estimates of the response function from each pair of images using a Kalman filter [22]. The state is the coefficients of the response function ($\phi = [\mathbf{c_1}, \ldots, \mathbf{c_M}]^\mathbf{T}$) and it is assumed to remain constant. Hence the process noise covariance was set to zero and the time update equations used are

$$\begin{aligned} \widehat{\phi}_k^- &= \widehat{\phi}_{k-1} \\ \mathbf{P}_k^- &= \mathbf{P}_{k-1} \end{aligned} \tag{47}$$

where $\widehat{\phi}$ is the estimate of the state and $\mathbf{P}$ is the estimate error covariance matrix. The measurement update equations are

$$\begin{aligned} \kappa_k &= \mathbf{P}_k^-(\mathbf{P}_k^- + \mathbf{R})^{-1} \\ \widehat{\phi}_k &= \widehat{\phi}_k^- + \kappa_k(\mathbf{z}_k - \widehat{\phi}_k^-) \\ \mathbf{P}_k &= (\mathbf{I} - \kappa_k)\mathbf{P}_k^- \end{aligned} \tag{48}$$

where κ is the Kalman gain, $\mathbf{z}_k$ is the measurement which is the pair-wise estimate of the response function in our case, and $\mathbf{R}$ is the measurement noise covariance. Let $\mathbf{D} = (-\mathbf{W}^T\mathbf{U}^{-1}\mathbf{W} + \Lambda)$ and $\mathbf{b} = -\mathbf{W}^T\mathbf{U}^{-1}\mathbf{v} + \mathbf{m}$ from Eq. (40), the covariance matrix $\mathbf{R}$ is computed as follows.

$$\mathbf{R} = (\mathbf{D}^T\mathbf{D})^{-1}((\mathbf{D}\upsilon - \mathbf{b})^T(\mathbf{D}\upsilon - \mathbf{b})) \tag{49}$$

The Kalman estimate of the response function $\widehat{\phi} = [\hat{c_1}, \ldots, \hat{c_M}]^T$ is incorpo-

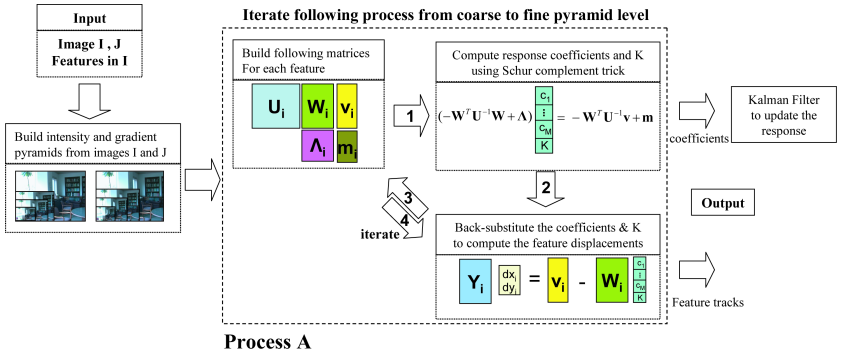

Fig. 4. Overview of our algorithm

rated to the response function estimation in the next frame in the sequence where the problem becomes estimating Δc_k as follows.

$$g(I) = \mathbf{g}_0(I) + \sum_{k=1}^{M} (\hat{c_k} + \Delta c_k)\mathbf{h}_k(I) \tag{50}$$

3.4 Multi-scale Iterative Algorithm

Fig. 4 shows the overview of our algorithm for the method explained in Sec. 3.2. As with the standard KLT tracker implementation, our algorithm runs iteratively on multiple scales. Image intensity and gradient pyramids are first built and the computation (process A in Fig. 4) starts from the coarsest level pyramid to the finest level. The process A in Fig. 4 is iterated multiple times for each pyramid level. The output of the algorithm are the coefficients for the response function which are fed to the Kalman filter (Sec. 3.3), the exposure difference K, and the tracked features which become input for the next pair of frames. Notice that we can start the tracking process with unknown response function and switch to tracking with known response function explained in section 3.1 when the estimate of the response function gets stable.

4 Stereo

Now that the images are radiometrically calibrated, and feature tracks have been recovered, the next step in our system is to perform structure from motion and stereo to recover a dense 3D surface.

Using the feature tracks, we recover the camera motion using the techniques found in [19]. In our experiments, we calibrated the camera intrinsics manually, although auto-calibration could have been used instead. The ability to track features in spite of exposure variations is essential for camera tracking in high dynamic range scenes such as the outdoors. Using our method, we are able to recover camera motion despite passing in and out of heavy shadows and even entering fully enclosed areas. Furthermore, feature tracks are continued over a larger number of frames, which is important for reducing drift in bundle adjustment.

Because our tracker also recovers the radiometric calibration, our stereo method requires no special invariance to brightness changes. This allows us to use the simple sum of absolute differences (SAD) as a matching metric. We have extended our previous stereo work found in [11], which accounted only for changes in exposure (gain), to take into account the recovered radiometric response function. Thus our matching function for a given pixel (x, y) and disparity (depth) d is as follows.

$$C(x, y, d) = \sum_{i,j \in W} |g(I(x + i, y + j))) - g(J(x + i - d, y + j)) - K| \qquad (51)$$

where g is the recovered response function and K is the recovered exposure change between images I and J (Eq. (5)). The cost function is aggregated

Fig. 5. Feature tracking result (synthetic example) using : (first) standard KLT (second) local-adaptive KLT (third) our method with known response (fourth) our method with unknown response. Images are from [2]

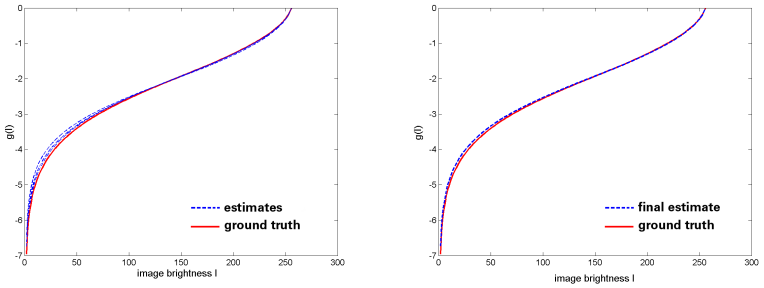

Fig. 6. (First) Samples of response functions estimated from the synthetic sequence. (Second) Final estimate of the response function.

over a window W, and the disparity (depth) with minimum cost is selected. As in our previous work, we use a planesweeping approach to handle multiple views simultaneously. This approach is both robust and efficient, and can be implemented on the GPU for real-time performance. For more details we refer the reader to our previous work [11].

5 Experimental results

We first evaluate our proposed methods with synthetic examples using evaluation images from [2]. The brightness of an image can be changed from I to I' using Eq. (52) with a response function g together with an exposure difference of K.

$$I' = g^{-1}(g(I) + K) \quad . \tag{52}$$

The response function used for the evaluation with the synthetic data is shown in Fig. 6. The exposure value applied for the examples from Fig. 5 was 0.4. The feature tracking results using the standard KLT ([15,20]), the local adaptive KLT ([2]), our method with known response function (Sec. 3.1), and our method with unknown response function (Sec. 3.2) are shown in Fig. 5. As expected, the standard KLT does not perform well under the brightness change. Our experiments show that the local adaptive KLT mostly performs well when the camera motion and the brightness change are small. However, the performance significantly degrades when the change in motion or brightness increases as demonstrated in this example. Tracking results using our methods, both with and without the knowledge of the response function, show superior results even with significant change in brightness which poses some problems for other tracking methods. The exposure value computed by our method was 0.404 with the known response function method and 0.408 with the unknown response function method. We further tested our response function estimation algorithm by creating a synthetic sequence with 9 images with varying exposure values. Fig. 6 shows some samples of the successive response function estimates and the final estimate along with the ground truth. Some estimates are less accurate in the lower intensity regions because the exposure difference was small in those image pairs. When the exposure difference is small, there are no changes in the brightness in the lower brightness regions giving no constraints to the estimation problem.

Similar results were observed in an experiment with a real video sequence. It was taken in a high dynamic range scene with a Canon GL2 camera. The exposure was automatically adjusted to a high value when the camera pointed to the dark inside area and it changed to a low value as the camera turned to the bright outside area. The comparison of tracks using the local-adaptive

Fig. 7. Feature Tracking using (First) Local-adaptive KLT (Second) Our method with known response (Third) Our method with unknown response. The video can be seen at http://www.cs.unc.edu/∼sjkim/klt/tracks.wmv

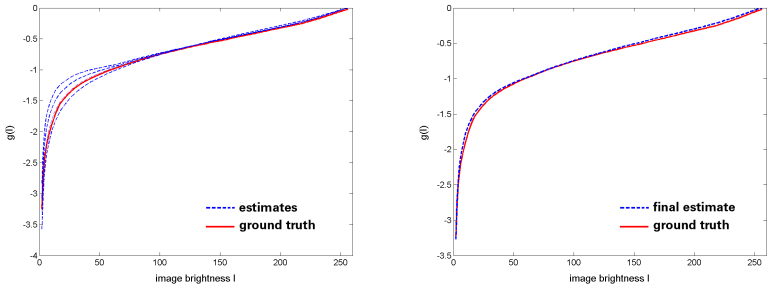

Fig. 8. (First) Samples of response functions estimated from the real video sequence (20 frames). (Second) Final estimate of the response function. The video can be seen at http://www.cs.unc.edu/∼sjkim/klt/track-response.wmv

KLT, our method with known response function, and our method with unknown response function is shown in Fig. 7. Both of our methods are able to track more features with significantly less errors when the change in motion and brightness is relatively large as shown in the example.

Fig. 8 shows the result of our response function estimation from this video. For the ground truth, we took multiple images of a static scene with a fixed camera changing the exposure value and fit the empirical model of response (EMoR) to the data as the method in [12]. Samples of the response function estimates and the final estimate are compared with the ground truth in Fig. 8.

We further verify our exposure estimation by comparing our estimates with the ground truth. Using a Point Grey Flea camera which has a linear response function, we took a video of a scene where the camera goes in and out of shadows causing the exposure to change frequently as shown in Fig. 9.

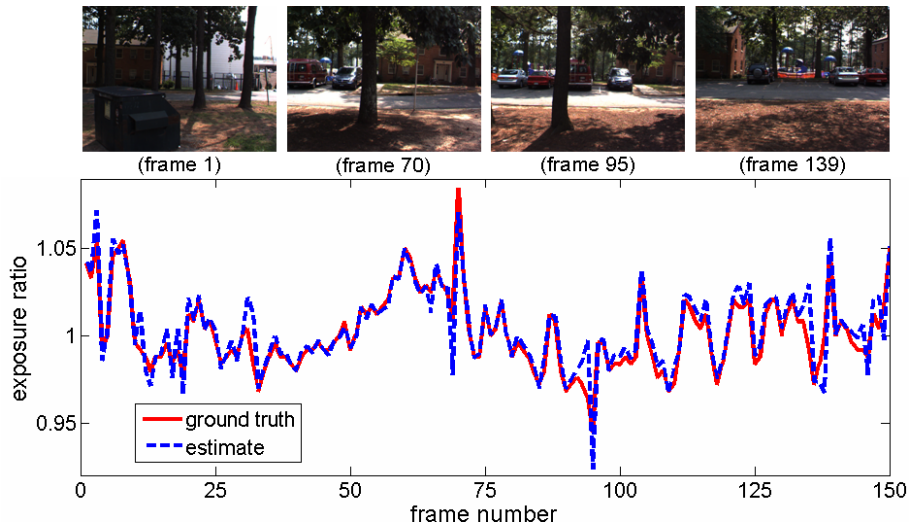

Fig. 9. Exposure Estimation. The estimated exposure values are compared with the values reported by the camera.

The computed exposure values are compared with the values reported by the camera in Fig. 9.

The execution time for tracking 500 features in 720x480 images on a Pentium 4 processor (2.80 GHz) was 5 frames/second for the standard KLT, the local-adaptive KLT, and our method with known response. For our method with unknown response, the execution time was 0.2 frames/second which includes camera response and exposure estimation in addition to tracking. Only a few frames are necessary to compute the response function and our method with the known response can be used for tracking afterwards. The overhead would be about 5% to 10% when tracking a 1-minute video.

To evaluate our stereo algorithm, we used videos from two scenes with high dynamic range which causes the exposure to change significantly. Some sample images of the videos are shown in Fig. 10 and Fig. 11. For the first example, the exposure changes because the camera moves from a shadow to sunlight. Depth map computed with our method which adapts for the exposure change shows superior results compared with the depth map computed

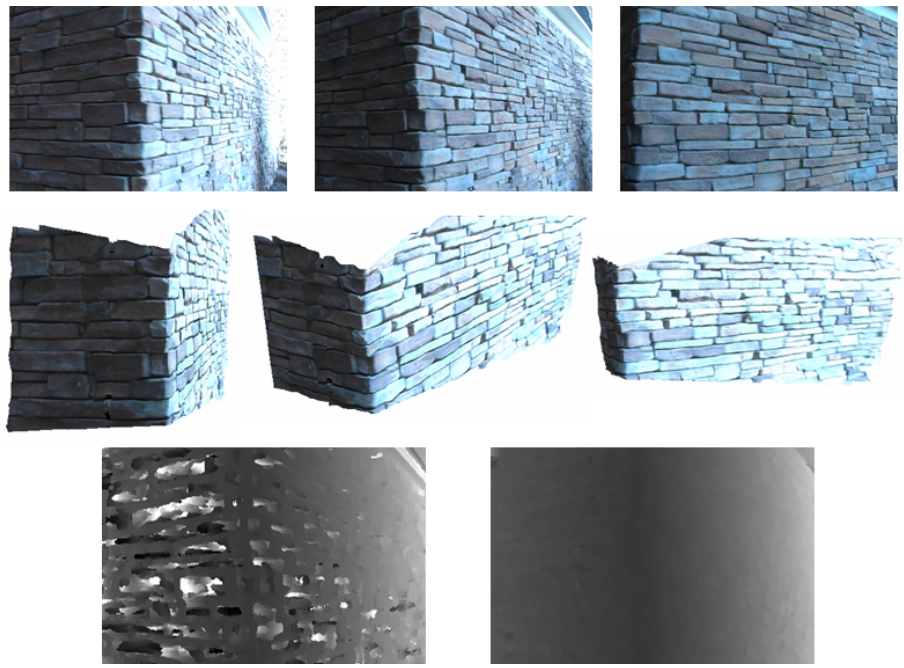

Fig. 10. First stereo example. (Top) Sample images from the video sequence (Middle) Generated 3D model with radiometrically aligned textures (Bottom) Depth maps computed without exposure compensation (left) and with our method (right)

without compensating for the exposure change as shown in the last row of Fig. 10. The texture-mapped 3D model of the scene generated with our stereo system are also shown in Fig. 10. The textures are radiometrically aligned to a constant exposure values using Eq. 52. For the second stereo example, a video of a tunnel-like structure is taken starting from outside. This example is more challenging due to bigger exposure changes and more complex geometry of the scene. Depth map comparison and 3D models in Fig. 11 show similar result as the first example. Note that the some textures are radiometrically distorted in the model because the original pixels for those regions were saturated.

6 Conclusion

We have introduced a stereo system that adapts to exposure changes by developing a novel method that unifies the problems of feature tracking and

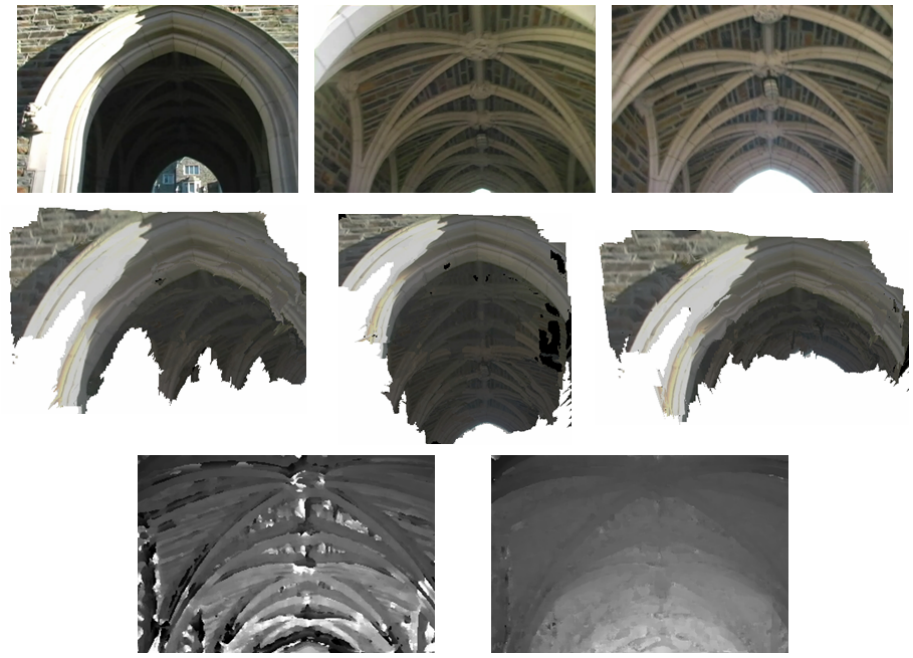

Fig. 11. Second stereo example. (Top) Sample images from the video sequence (Middle) Generated 3D model with radiometrically aligned textures (Bottom) Depth maps computed without exposure compensation (left) and with our method (right)

radiometric calibration into a common framework. For feature tracking, it is commonly required that the brightness of features stay constant or the variations are dealt locally and linearly when the change is actually global and nonlinear. This limitation is not acceptable in many applications like building 3D models from outdoor scene videos in which high dynamic range environments are captured with a low dynamic range camera system. To overcome these limitations, we proposed a joint feature tracking, radiometric response function and exposure estimation framework. This solves the chicken-and-egg problem in which the tracking requires accurate radiometric calibration for accuracy which in turn relies on precise tracks. Our computationally efficient algorithm takes advantage of the structure of the estimation problem which leads to a minimal computational overhead. With our joint estimation, we were able to advance the quality and robustness of the known structure from motion techniques [19] by incorporating the information for 3D camera tracking, the depth from stereo and providing radiometrically aligned images for

texture-mapping. In the future, we plan to add vignetting estimation to the process by using the tracks over multiple frames. In addition, we will also explore the possibility of applying our method for creating high dynamic range (HDR) videos [9].

References

[1] S. Baker, R. Gross, I. Matthews, T. Ishikawa, Lucas-kanade 20 years on: A unifying framework: Part 2, Tech. Rep. CMU-RI-TR-03-01, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA (February 2003).

[2] S. Birchfield, Klt: An implementation of the kanade-lucas-tomasi feature tracker, http://www.ces.clemson.edu/ stb/klt/.

[3] F. Candocia, Jointly registering images in domain and range by piecewise linear comparametric analysis, IEEE Transactions on Image Processing (2003) 409–419.

[4] P. Debevec, J. Malik, Recovering high dynamic range radiance maps from photographs, Proc. SIGGRAPH'97 (1997) 369–378.

[5] M. Grossberg, S. Nayar, Determining the camera response from images: What is knowable?, IEEE Transaction on Pattern Analysis and Machine Intelligence 25 (11) (2003) 1455–1467.

[6] M. Grossberg, S. Nayar, Modeling the space of camera response functions, IEEE Transaction on Pattern Analysis and Machine Intelligence 26 (10) (2004) 1272–1282.

[7] H. Hirschmuller, D. Scharstein, Evaluation of cost functions for stereo matching, Proc. IEEE Conference on Computer Vision and Pattern Recognition (2007) 1–8.

[8] H. Jin, P. Favaro, S. Soatto, Real-time feature tracking and outlier rejection with changes in illumination, Proc. IEEE Int. Conf. on Computer Vision (2001) 684–689.

[9] S. B. Kang, M. Uyttendaele, S. Winder, R. Szeliski, High dynamic range video, ACM Transactions on Graphics 22 (3) (2003) 319–325.

[10] S. J. Kim, J.-M. Frahm, M. Pollefeys, Joint feature tracking and radiometric calibration, Proc. IEEE Int. Conf. on Computer Vision (2007) 1–8.

[11] S. J. Kim, J.-M. Frahm, M. Pollefeys, S. j. kim and d. gallup and j.-m. frahm and a. akbarzadeh and q. yang and r. yang and d. nistr and m. pollefeys, Proc. Int. Conf. on Computer Vision Systems.

[12] S. J. Kim, M. Pollefeys, Radiometric alignment of image sequences, Proc. IEEE Conference on Computer Vision and Pattern Recognition (2004) 645–651.

[13] S. Lin, J. Gu, S. Yamazaki, H. Shum, Radiometric calibration from a single image, Proc. IEEE Conference on Computer Vision and Pattern Recognition (2004) 938–945.

[14] S. Lin, L. Zhang, Determining the radiometric response function from a single grayscale image, Proc. IEEE Conference on Computer Vision and Pattern Recognition (2005) 66–73.

[15] B. D. Lucas, T. Kanade, An iterative image registration technique with an application to stereo vision, Conference on Artificial Intelligence (1981) 674–659.

[16] S. Mann, 'pencigraphy' with agc: Joint parameter estimation in both domain and range of functions in same orbit of the projective-wyckoff group, Proc. IEEE International Conference on Image Processing (1996) 193–196.

[17] S. Mann, R. Picard, On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures, Proc. IS&T 46th

annual conference (1995) 422–428.

[18] T. Mitsunaga, S. Nayar, Radiometric self-calibration, Proc. IEEE Conference on Computer Vision and Pattern Recognition (1999) 374–380.

[19] M. Pollefeys, L. V. Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops, R. Koch, Visual modeling with a hand-held camera, International Journal of Computer Vision 59 (3) (2004) 207–232.

[20] J. Shi, C. Tomasi, Good features to track, Proc. IEEE Conference on Computer Vision and Pattern Recognition (1994) 593–600.

[21] B. Triggs, P. McLauchlan, R. Hartley, A. Fitzgibbon, Bundle adjustment – a modern synthesis, in: B. Triggs, A. Zisserman, R. Szeliski (eds.), Vision Algorithms: Theory and Practice, vol. 1883 of Lecture Notes in Computer Science, Springer-Verlag, 2000.

[22] G. Welch, G. Bishop, An introduction to the kalman filter, Tech. Rep. TR-95-031, Department of Computer Science, University of North Carolina at Chapel Hill (December 1995).