A Model of Brightness Variations Due to Illumination Changes and Non-rigid Motion Using Spherical Harmonics

José M. Buenaposada
Dep. Ciencias de la Computación,
U. Rey Juan Carlos, Spain
`http://www.dia.fi.upm.es/~pcr`

Alessio Del Bue
Inst. for Systems and Robotics
Inst. Superior Técnico, Portugal
`http://www.isr.ist.utl.pt/~adb`

Enrique Muñoz
Facultad de Informática,
U. Complutense de Madrid, Spain
`http://www.dia.fi.upm.es/~pcr`

Luis Baumela
Dep. de Inteligencia Artificial,
U. Politécnica de Madrid, Spain
`http://www.dia.fi.upm.es/~pcr`

Abstract

Pixel brightness variations in an image sequence depend both on the objects' surface reflectance and on the motion of the camera and object. In the case of rigid shapes some proposed models have been very successful explaining the relation among these strongly coupled components. On the other hand, shapes which deform pose new challenges since the relation between pixel brightness variation with non-rigid motion is not yet clear. In this paper, we introduce a new model which describes brightness variations with two independent components represented as linear basis shapes. Lighting influence is represented in terms of Spherical Harmonics and non-rigid motion as a linear model which represents image coordinates displacement. We then propose an efficient procedure for the estimation of this image model in two distinct steps. First, shape normals and albedo are estimated using standard photometric stereo on a sequence with varying lighting and no deformable motion. Then, given the knowledge of the object's shape normals and albedo, we efficiently compute the 2D coordinates bases by minimizing image pixel residuals over an image sequence with constant lighting and only non-rigid motion. Experiments on real tests show the effectiveness of our approach in a face modelling context.

1 Introduction

Being able to deal with non-rigid motion is a Computer Vision problem that has received a lot of attention in recent years. The main reason is that there are very interesting deformable objects, from the application point of view, like the face. In order to deal with non-rigid objects we need a model that is accurate enough to precisely describe and separate all sources of variation in object's appearance. To do this we can build off-line, and

possibly automatically, a model able to represent the object under any imaging condition. These models are a description of certain properties such as the 3D shape, surface normals, the image texture and the reflectance properties of object's surface. All these models can be used for different tasks such as object tracking, detection or recognition.

The ongoing trend of the last years is to compute the object models in a generative manner – each model is estimated from a training set of images and not tuned from a priori modelling of the data. Thus, in Structure from Motion (SfM) trajectories of points from a set of sparse features are used to extract rigid [15] and non-rigid [4] 3D models. In Photometric Stereo (PS) [1] a set of images of a rigid shape under different lighting conditions is used to compute surface normals (a needle map) and albedo. In Active Appearance Models (AAM) [6, 12] the image variations in a video are used to obtain a texture parametrization together with a set of 2D deformable basis.

In recent years the relation between illumination changes, shape of the object and subspace models of illumination (built from a PCA over a set of illumination varying sequence of images) has been explained in terms of Spherical Harmonics [2]. Even the case of changing lighting and rigid motion of the objects has been successfully modeled [18, 17]. Anyhow, there is not such a model explaining simultaneously the influence of illumination variations and non-rigid motion on the appearance of a deformable object.

In this paper we first show that, with linear assumptions, it is possible to model the influence of non-rigid motion and illumination in appearance as a set of two independent basis. Our work is based on the description of the reflectance properties of a surface as a set of Spherical Harmonics basis [2]. Then, we join this model with the 2D modes of deformation at each pixel given by the non-rigid motion and find a complete description of the brightness variations with two independent linear subspaces of deformation and illumination. Therefore we propose an extension of previous approaches to the case of a static but deforming object when the illumination is varying.

Given this new theoretical model, we then define a procedure able to extract these bases as a two step algorithm. Initially we compute surface normals and albedo with a sequence where a static target is illuminated by a moving light source using the Basri and Jacobs's approach [1]. Next, given the surface normals and albedo, we compute the 2D modes of deformation from a sequence with the subject deforming under three distinct light sources. Given the first stage, this step reduce to a simple PCA computation over the training images followed by a LS estimate for each image pixel.

2 The joint illumination-deformation model

We introduce a new appearance-based model representing the variations in the image caused by changes both in 2D non-rigid motion and in the illumination of the scene. Our new model is based on a linear approximation to the object's reflectance function. The set of images of a convex object with Lambertian reflectance function forms a convex cone in the space of all possible images [3]. If we make the usual assumption that a human face is approximately a convex Lambertian object [7], then the illumination cone model is applicable to the face modelling problem. In the context of face recognition, and restricted to a single image, Basri and Jacobs [2] derived a linear approximation to the illumination cone introduced in [3]. Their derivation is based on spherical harmonics, which are the equivalent for the sphere to the Fourier basis on the line or on the plane.

More recently, the problem of modelling the reflectance of a moving object in a video sequence has also been investigated. Yue et al. [18] studied the problem for a rotating face and proved that 2D harmonic basis images at different face poses are related by closed-form linear combinations. Simultaneously Xu and Roy-Chowdhury considered the more general problem of a translating and rotating object and proved that the set of images can be approximated by a bilinear subspace consisting of nine illumination and six motion variables [17]. In this section we extend the work in [18, 17] to the case of static but deforming objects. When rigid object motion is present, it will be pre-estimated using an image warping function.

2.1   First order approximation to a deforming object

Let $\mathtt{I}(\mathbf{x},t)$ be the matrix storing the image acquired at time t, where $\mathbf{x}$ is a vector representing the coordinates of a point in the image, and let $\mathbf{I}(t)$ be a vector storing the intensity values of $\mathtt{I}(\mathbf{x},t)$. The image $\mathbf{I}(t)$ can be linearly modelled with an infinite series of *harmonic images* [2],

$$\mathbf{I}(t) = \sum_{n} \sum_{m=-n}^{+n} l_{nm}(t)\mathbf{b}_{nm}(t), \tag{1}$$

where $\mathbf{b}_{nm}$ are the harmonic images and l_{nm} are the illumination coefficients encoding the direction and intensity of illumination. Let $\mathbf{x}_i = (x_i, y_i)$ be the image coordinates of pixel i. The harmonic image [2] for pixel i can be expressed in terms of $\mathbf{n}(\mathbf{x}_i, t)$, the normal to the surface of the target at pixel i on time t,

$$b_{nm}(i,t) = \rho(\mathbf{x}_i, t)\alpha_n Y_{nm}(\mathbf{n}(\mathbf{x}_i, t)), \tag{2}$$

where $\rho(\mathbf{x}_i, t)$ is the albedo of the 3D point that projects on pixel i at time t, α_n is a coefficient constant for each harmonic of order n and Y_{nm} is the surface spherical harmonic of order n.

The vector $\mathbf{I}(0)$ is the reference image of our object and it is chosen as an image with an arbitrary illumination and deformation. It represents the origin of the reference system generated by the model illumination and deformation parameters. Let $\mathbf{I}(t)$ be another image acquired at time t. We will show that a first order approximation to $\mathbf{I}(t)$ can be obtained by adding a linear combination of a motion and an illumination linear subspace to the reference template $\mathbf{I}(0)$. The albedo $\rho(\mathbf{x}_i, t)$ can be linearly approximated as:

$$\rho(\mathbf{x}_i, t) = \rho(\mathbf{x}_i, 0) + \underbrace{\left[\left. \frac{d\rho(\mathbf{x}, 0)}{d\mathbf{x}} \right|_{\mathbf{x}=\mathbf{x}_i} \right]^\top}_{\nabla_x \rho(\mathbf{x}_i, 0)} \underbrace{\left[\left. \frac{d\mathbf{x}_i}{dt} \right|_{t=0} \right]}_{\dot{\mathbf{x}}_i} \delta t + o(\delta t^2), \tag{3}$$

where $\dot{\mathbf{x}}_i$ is the 2D coordinates change corresponding to each pixel i on the reference image, due to non-rigid motion at any given time t. Assuming the changes in the normals due to deformations at each point are small the spherical harmonic function $Y_{nm}(\mathbf{n}(\mathbf{x}_i, t))$, may also be linearly approximated as

$$Y_{nm}(\mathbf{n}(\mathbf{x}_i, t)) = Y_{nm}(\mathbf{n}(\mathbf{x}_i, 0)) + \underbrace{\left[\left. \frac{dY_{nm}(\mathbf{n})}{d\mathbf{n}} \right|_{\mathbf{n}=\mathbf{n}(\mathbf{x}_i, 0)} \right]^\top}_{\nabla_{\mathbf{n}} Y_{nm}(\mathbf{n}(\mathbf{x}_i, 0))} \underbrace{\left[\left. \frac{d\mathbf{n}(\mathbf{x}, 0)}{d\mathbf{x}} \right|_{\mathbf{x}=\mathbf{x}_i} \right]}_{\nabla_x \mathbf{n}(\mathbf{x}_i, 0)} \dot{\mathbf{x}}_i \delta t + o(\delta t^2).$$

$$\tag{4}$$

This assumption over the normals may introduce estimation errors whenever shape areas with high curvature change due to deformations, however, considering weak variations in the normals, itf is generally reasonable for smooth deforming shapes. Now, substituting equations (3) and (4) into (2) and ignoring higher order terms we get a first order approximation to $b_{nm}(i,t)$ such that:

$$b_{nm}(i,t) = \underbrace{\rho(\mathbf{x}_i,0)\alpha_n Y_{nm}(\mathbf{n}(\mathbf{x}_i,0))}_{b_{nm}(i,0)} + \mathbf{a}_{nm}^\top(i)\dot{\mathbf{x}}_i\delta t + o(\delta t^2), \tag{5}$$

where $\mathbf{a}_{nm}^\top(i)$ is the 1×2 vector modelling the change of albedo and spherical harmonics (*deformation brightness change*) at pixel i caused by the deformation of the target and it is equal to:

$$\mathbf{a}_{nm}^\top(i) = \rho(\mathbf{x}_i,0)\alpha_n \nabla_{\mathbf{n}} Y_{nm}(\mathbf{n}(\mathbf{x}_i,0))\nabla_x \mathbf{n}(\mathbf{x}_i,0) + \nabla_x \rho(\mathbf{x}_i,0)\alpha_n Y_{nm}(\mathbf{n}(\mathbf{x}_i,0)). \tag{6}$$

The first order approximation to the illumination coefficients may also be expressed as

$$l_{nm}(t) = l_{nm}(0) + \underbrace{\left[\frac{dl_{nm}(t)}{dt}\bigg|_{t=0}\right]}_{\dot{l}_{nm}(0)} \delta t + o(dt^2). \tag{7}$$

From equations (1), (5), (7) and ignoring higher order terms we get a first-order approximation to the intensity value of pixel i at time t

$$\begin{aligned} I(i,t) &= \sum_n \sum_{m=-n}^{+n} [l_{nm}(0)b_{nm}(i,0)] + \sum_n \sum_{m=-n}^{+n} [b_{nm}(i,0)\dot{l}_{nm}(0)\delta t] + \\ &\quad \sum_n \sum_{m=-n}^{+n} \left[l_{nm}(0)\mathbf{a}_{nm}^\top(i)\dot{\mathbf{x}}_i\delta t\right] + o(\delta t^2) \end{aligned} \tag{8}$$

Now we have to express analytically the image displacement $\dot{\mathbf{x}}_{i,0}$ at each frame. In our case, since we assume no rigid motion, we have that every pixel displacement is given only by deformations. This 2D motion is parameterised as a linear combination of basis shapes $\mathbf{S}_i$ centered at the initial pixel position $\mathbf{x}_{i,0}$ such that $\mathbf{x}_{i,t} = \mathbf{x}_{i,0} + \mathbf{S}_i\mathbf{c}_d(t)$ where each $\mathbf{S}_i$ is a $2 \times d$ matrix and $\mathbf{c}_d$ is a $d \times 1$ vector of deformation coefficients [1]. Given this model, we can relate it to the displacement $\dot{\mathbf{x}}_{i,0}$ in equation (8) such that

$$\dot{\mathbf{x}}_{i,0}\delta t = \mathbf{x}_{i,t} - \mathbf{x}_{i,0} + o(dt^2) = \mathbf{S}_i\mathbf{c}_d(t) + o(dt^2) \tag{9}$$

represents a first order approximation to deformations. Similarly, from equation (7), we approximate the variation of the illumination coefficients as:

$$\dot{l}_{mn}(0)\delta t = l_{mn}(t) - l_{mn}(0) + o(dt^2) = \Delta l_{mn}(t) + o(dt^2) \tag{10}$$

Finally, substituting equations (9) and (10) into (8) we obtain

$$\mathbf{I}(t) = \sum_n \sum_{m=-n}^{+n} [l_{nm}(0)\mathbf{b}_{nm}(0) + \Delta l_{nm}(t)\mathbf{b}_{nm}(0) + l_{nm}(0)\mathbf{A}_{nm}\mathbf{c}_d(t)] + o(\delta t^2),$$

---

[1]Notice that the matrix $\mathbf{S}_i$ represents 2D modes of deformation as similarly expressed by AAMs [6, 12]
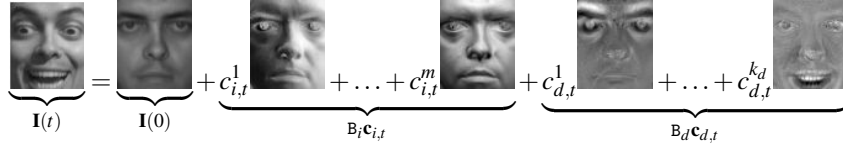
Figure 1: A pictorial representation of equation (11)

where

$$A_{nm} = \begin{bmatrix} \mathbf{a}_{nm}^{\top}(1)S_1 \\ \vdots \\ \mathbf{a}_{nm}^{\top}(N)S_N \end{bmatrix}, \quad N = \text{size}(\mathbf{I}(t)),$$

which can be rewritten in a compact matrix form as

$$\mathbf{I}(t) = \mathbf{I}(0) + B_i\mathbf{c}_{i,t} + B_d\mathbf{c}_{d,t}. \tag{11}$$

The first term in equation (11) refers to the reference image, $\mathbf{I}(0) = \sum_n \sum_{m=-n}^{+n} l_{nm}(0)\mathbf{b}_{nm}(0)$. The second term models the appearance change of the reference image caused by a variation in illumination. The vector $\mathbf{c}_{i,t} = [\Delta l_{0,0}(t), \Delta l_{1,-1}(t), \Delta l_{1,0}(t), \ldots]^{\top}$ encodes the change of direction and intensity of illumination at time t, whereas the matrix $B_i = [\mathbf{b}_{0,0}(0), \mathbf{b}_{1,-1}(0), \mathbf{b}_{1,0}(0), \ldots]$ represents the illumination basis, which is composed by the harmonic reference images. Finally, the third term represents the change of target appearance caused by a deformation. The vector $\mathbf{c}_{d,t}$ stores the target shape coefficients and the matrix $B_d = \sum_n \sum_{m=-n}^{+n} l_{nm}(0)A_{nm}$ models the image changes caused by the deformation of the target. Notice that equation (11) is an expression of the brightness constancy equation in the case of a varying illumination and deforming object (see Figure 1).

The dimension of the illumination model is related to the order of the harmonic approximation. A second order harmonic approximation suffices to represent 97.96% of the energy in $\mathbf{I}(t)$ [2]. Therefore, if we only consider a second order approximation ($n = 0, 1, 2$), the dimension of matrix B_i will be at most nine. On the other hand, B_d is a set of bases which depends on the modes of deformation and its dimensionality k_d increases with the degrees of non-rigidity of the shape.

Linear models as represented in equation (11) have been used previously in the context of illumination invariant 3D and 2D rigid face tracking. LaCascia et al. [11] introduced a 3D rigid face model, whose constancy equation also had two independent linear subspaces. Each subspace represented illumination and rigid motion warping templates respectively. Similarly, Hager and Belhumeur [9] used a 2D rigid face model with a similar constancy equation in which a single linear subspace was used to model changes in illumination. Our model differs from these since the image shape is allowed to deform and the appearance model is computed in the image plane, rather than the texture map plane, as is the case of [11]. Like our model, Tenenbaum and Freeman's [14] and Grimes et al.'s [8] bilinear models and Vasilescu and Terzopoulos' multilinear model [16] also represented variations in illumination and facial expression, but illumination and appearance are not independent in their models. Buenaposada et al. [5] presented a similar model assuming illumination and face deformation subspaces are independent but without giving a theoretical explanation only giving experimental evidence on the face tracking problem.

Figure 2: (a) Images used to perform the photometric stereo reconstruction [1]. (b) The 3D surface obtained by integrating the normals at each pixel.

3 Non-rigid motion estimation

The linear approximations introduced in Section 2 leads to the formulation of the brightness constancy equation (11) in which the grey level image variations are a function both of deformation and lighting changes. In such way, linear basis components computed using PCA from image observations of a deforming body in varying illumination conditions shows a dimensionality which depends both on the 2D deformable basis B_d and the spherical harmonics basis B_i. However now the complexity resides in understanding which component from PCA is associated to the varying illumination conditions and which one to the deformations. Since these components are collapsed into a single pixel brightness varying through time, such disambiguation may result in a hard problem. On the other hand, taking for granted that PCA may directly separate the components as given in equation (11) is an unrealistic assumption.

Nonetheless there are two evident cases when this problem can be solved. First, when the image shape is under the effect of varying illumination while the light source is changing direction and position (photometric stereo, $B_d \mathbf{c}_{d,t} = costant$). Second, when the light source is fixed and the shape is deforming (2D motion estimation, $B_i \mathbf{c}_{i,t} = costant$). If two image sequences are given in such constraints, it is then possible to first estimate the basis B_i (computing shape normals and albedo) and then to compute the deforming flow basis B_d given the estimated lighting parameters.

3.1 Photometric Stereo

In order to fully characterise the models expressed in equation (11), we first compute shape normals $\mathbf{n}(\mathbf{x}_i, 0)$ and albedo $\rho(\mathbf{x}_i, 0)$ using an image sequence showing a rigid configuration with varying lighting conditions. The components $\mathbf{n}(\mathbf{x}_i, 0)$ and $\rho(\mathbf{x}_i, 0)$ will then be used to obtain the luminance variations basis as in equation (2) and the $\mathbf{a}_{nm}(i)$ *deformation brightness change* vector as in equation (6) which are needed for the following deformable bases estimation. The computation of the normals and the albedo is a standard photometric stereo problem that can be solved in closed-form and without prior calibration using Basri and Jacob's algorithm [1]. Our choice is driven towards this approach since we want to maintain a calibration-free formulation of our technique. Figure 2 shows the image sequence of a face under varying lighting used for one of the test in the experimental section. The computed 3D shape is given by the integration of the normal components and it shows the correctness of the reconstruction.

3.2 A solution for the 2D deformation bases

Once the *deformation brightness change* vector $\mathbf{a}_{nm}(i)$ and illumination subspace basis matrix B_i have been estimated, it is now possible to follow with the computation of the 2D deformation basis S_i. If we have purely deforming motion with fixed lighting condition, it follows that $\mathbf{c}_{i,t} = constant$ since image variations are given exclusively by shape motion. Thus equation (11) reduces to:

$$\mathbf{I}(t) = \mathbf{I}(0) + \mathrm{B}_d \mathbf{c}_{d,t}. \tag{12}$$

In order to estimate $\mathbf{I}(0)$ we have to choose an image on the deformation training sequence with the same object configuration (e.g. a neutral face expression). By projecting this image over the illumination subspace given by B_i we will have the $\mathbf{l}_{nm}(0)$ illumination coefficients necessary to compute $\mathbf{I}(0)$. The image residual at each frame $\tilde{\mathbf{I}}(t) = \mathbf{I}(t) - \mathbf{I}(0)$ is then expressed as a set of linear bases using PCA analysis giving:

$$\tilde{\mathbf{I}}(t) = \tilde{\mathrm{B}}_d \tilde{\mathbf{c}}_{d,t}, \tag{13}$$

where $\tilde{\mathrm{B}}_d$ is the basis matrix resulting from PCA of the deformation training sequence and the weights $\tilde{\mathbf{c}}_{d,t}$ which express the linear combination of the basis shapes. However, from the previous photometric stereo reconstruction, we partially know the structure of B_d since it is composed by the given $\mathbf{a}_{nm}(i)$ and the unknown 2D flow basis S_i such that:

$$\mathrm{B}_d = \sum_n \sum_{m=-n}^{+n} l_{nm}(0) \mathrm{A}_{nm} = \sum_n \sum_{m=-n}^{+n} l_{nm}(0) \begin{bmatrix} \mathbf{a}_{nm}^\top(1)\mathrm{S}_1 \\ \vdots \\ \mathbf{a}_{nm}^\top(N)\mathrm{S}_N \end{bmatrix}. \tag{14}$$

Now given $\mathbf{a}_{nm}(i)$ and the estimated PCA weights $\tilde{\mathbf{c}}_{d,t}$ we can build a system of linear equations to solve for the unknown deformation basis $\mathrm{S}_1 \dots \mathrm{S}_N$. For each frame t we can build a single equation given by:

$$\sum_n \sum_{m=-n}^{+n} l_{nm}(0) \tilde{\mathbf{c}}_{d,t}^T \otimes \mathbf{a}_{nm}^\top(i)^T \mathbf{s}_i = \tilde{\mathbf{I}}(t) \tag{15}$$

where $\mathbf{s}_i$ is the vectorization of the $2 \times k_d$ basis S_i (i.e. $\mathbf{s}_i = vec(\mathrm{S}_i)$) and $\otimes$ represents the Kronecker product. By stacking row-wise all the equations for each frame we obtain the overall system:

$$\mathrm{H}\mathbf{s}_i = \tilde{\mathbf{I}}_i \tag{16}$$

where H is a $F \times 2k$ matrix and $\tilde{\mathbf{I}}_i$ a F-vector that contains the image residuals for the image point i. The matrix H is rank deficient given the repetitive structure in (15). This shows that the flow is underconstrained for a single light source. However, under different lights sources we can build a stacked representation given each single light source contribution such that:

$$\bar{\mathrm{H}} = \begin{bmatrix} \mathrm{H}_1 \\ \vdots \\ \mathrm{H}_C \end{bmatrix} \tag{17}$$

where C is the number of light sources. The Least Squares solution for the system $\bar{\mathrm{H}}\mathbf{s}_i = \bar{\mathbf{I}}_i$ gives the flow basis for pixel i. Now $\bar{\mathbf{I}}_i$ is a (CF)-vector containing the image residuals
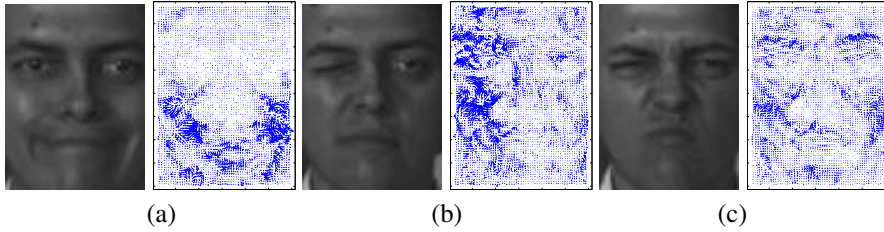
(a)   (b)   (c)

Figure 3: Image displacement estimated from different images. (a) An image from a 64 frames sequence with a smile expression ($k_d = 3$). (b) An image from a 29 frames sequence with one eye blinking ($k_d = 3$). (b) An image from a 48 frames sequence with two disgusted expressions ($k_d = 4$).

for each light. As we are dealing with deforming objects we need exactly the same shape configuration for each light on every training image. The only way of assuring this is to use a red, a green and a blue filtered lights and RGB images [10]. Each color channel of the acquired images will be a grey level image with different lighting and exactly the same shape deformation.

In order to increase estimation robustness, we compute the basis flow around an image patch W centered on the image pixel i. A positive function, that monotonically declines with the distance from the patch center, is then used to weigh each pixel in the subwindow. At the end, the final cost being minimised for each basis is the following sum of LS terms:

$$\arg\min_{\mathbf{s}_i} \sum_{p \in W} w_p ||\mathrm{H}_p \mathbf{s}_i - \bar{\mathbf{I}}_{ip}||_2^2 \qquad (18)$$

where w_p is the weight at each pixel p belonging to the subwindow W, H_p are derived from equation (17) at each pixel p and $\bar{\mathbf{I}}_{ip}$ the relative image residuals.

4   Experiments

Our approach was tested mainly in the case of image faces, with the shape performing different expression. Our aim was to obtain a description of the varying objects in terms of both luminance and deformation basis. The first stage using the photometric stereo [1] computed both the albedo and normals. As already seen in Figure 2 a reasonable estimation can be obtained with faces even with a small images (the image size for all the frames in this section is 61×81).

After obtaining the luminance basis, a second sequence showing the face with different facial expressions was used to infer the deformation basis. Here, the subject was standing still in front of the camera and a rigid registration algorithm was used to compensate for small motions of the subject. The shape was illuminated by three colored lights, thus we obtained three image sequences for each channel. After performing PCA on the stacked vectorised images, we used the PCA weights and the $\mathbf{a}_{nm}(i)$ to build the linear system as in equation (18) (i.e. $C = 3$) and solve for the deformation basis. A subwindow W of 5×5 pixels was used.

We then follow with the computation of the 2D basis shapes by performing a PCA over the sequences with varying expressions under three different light sources. As it can
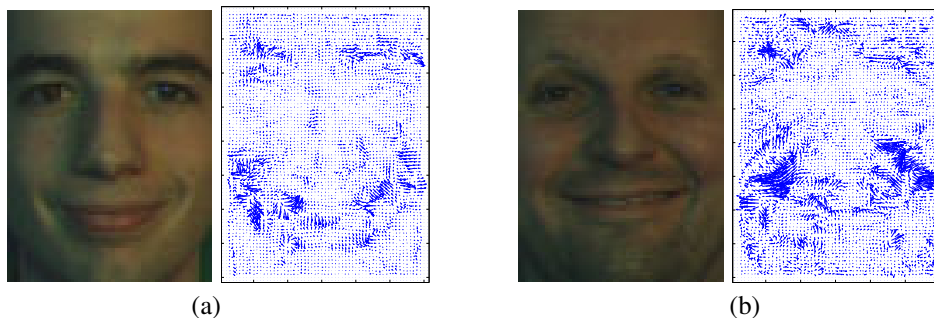
Figure 4: Image displacement estimated from different images. Both sequences have a smile expression. (a) A 101 frames sequence with ($k_d = 3$). (b) A 21 frames sequence with ($k_d = 2$).

be seen in Figure 3, at each expression corresponds localized image displacement with an orientation corresponding to the deformation motion in respect to a neutral face.

Figure 4 presents two further subjects reconstruction showing both a smile expression. Once again the displacement given the estimated 2D displacement is mostly coherent with the image motion. The dimension k_d of the PCA projection is determined using *parallel analysis* [13]. Basically, we compare the covariance matrix of the data with a randomised version of it to find the amount of eigenvalue inflation due to sampling errors.

5 Conclusions

Our approach is modelling image variations as a set of two independent bases describing illumination and 2D image coordinates variations. The presented theoretical model led to the definition of a two-step procedure to compute the bases from video sequences. In such way we obtain a compact parametrisation of the image shape in the illumination and in the shape deformation subspaces. As shown in the experiments, the model here introduced can be readily used for constraining the 2D image variations generated by a deforming object. Our future work also aims to extend this framework's theory in order to directly compute 3D deformation basis and to reduce the 3-color light setup to a single source. Another challenge remaining is to break the assumption of slight variation in the normals orientation. This would help modelling the shape areas with high curvature changes due to deformations.

References

[1] Ronen Basri, David Jacobs, and Ira Kemelmacher. Photometric stereo with general, unknown lighting. *International Journal of Computer Vision*, 72(3):239–257, 2007.

[2] Ronen Basri and David W. Jacobs. Lambertian reflectance and linear subspaces. *Trans. on PAMI*, 25(2):218–233, February 2003.

[3] Peter Belhumeur and David J. Kriegman. What is the set of images of an object under all possible illumination conditions? *International Journal of Computer Vision*, 28(3):245–260, 1998.

[4] C. Bregler, A. Hertzmann, and H. Biermann. Recovering non-rigid 3d shape from image streams. In *Proc. of CVPR*, pages 690–696, June 2000.

[5] J.M. Buenaposada, E. Muñoz, and L. Baumela. Efficient illumination independent appearance-based face tracking. *Image and Vision Computing*, (doi:10.1016/j.imavis.2008.04.015), 2008.

[6] T. Cootes, G.J. Edwards, and C. Taylor. Active appearance models. In *Proc. European Conference on Computer Vision*, volume LNCS 1047, pages 484–498. Springer-Verlag, 1998.

[7] Athinodoros S. Georghiades, Peter N. Belhumeur, and David J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *Trans. on PAMI*, 23(6):643–660, June 2001.

[8] D.B. Grimes, A.P. Shon, and R.P.N. Rao. Probabilistic bilinear models for appearance-based vision. In *Proc. of ICCV*, volume II, pages 1478–1485, 2003.

[9] Gregory Hager and Peter Belhumeur. Efficient region tracking with parametric models of geometry and illumination. *Trans. on PAMI*, 20(10):1025–1039, 1998.

[10] Carlos Hernández, George Vogiatzis, Grabriel J. Browstow, Bjorn Stenger, and Roberto Cipolla. Non-rigid photometric stereo with colored lights. In *Proc. of ICCV*, pages 1–8, 2007.

[11] Marco La Cascia, Stan Sclaroff, and Vasili Athitsos. Fast, reliable head tracking under varying illumination: An approach based on robust registration of texture-mapped 3d models. *Trans. on PAMI*, 22(4):322–336, April 2000.

[12] Iain Matthews and Simon Baker. Active appearance models revisited. *International Journal of Computer Vision*, 60(2):135–164, 2004.

[13] M. B. Stegmann. Analysis and segmentation of face images using point annotations and linear subspace techniques. Technical report, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, 2002.

[14] Joshua B. Tenenbaum and William T. Freeman. Separating style and content with bilinear models. *Neural Computation*, 12:1247–1283, 2000.

[15] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization approach. *International Journal of Computer Vision*, 9(2):137–154, 1992.

[16] M. Vasilescu and D. Terzopoulos. Multilinear analysis of image ensembles: Tensorfaces. In *Proc. European Conference on Computer Vision*, volume LNCS 2350, pages 447–460. Springer-Verlag, 2002.

[17] Yiley Xu and Amit K. Roy-Chowdhury. Integrating the effects of motion, illumination and structure in video sequences. In *Proc. of ICCV*, volume II, pages 1675–1682, 2005.

[18] Zhanfeng Yue, Wenyi Zhao, and Rama Chellappa. Pose-encoded spherical harmonics for robust face recognition using a single image. In Wenyi Zhao, Shaogang Gong, and Xiaoou Tang, editors, *Proc. Workshop on Analysis and Modelling of Faces and Gestures*, volume LNCS-3723. Springer-Verlag, 2005.