On Separation of Semitransparent Dynamic Images from Static Background

Alexander M. Bronstein[1], Michael M. Bronstein[1], and Michael Zibulevsky[2]

[1] Department of Computer Science, Technion – Israel Institute of Technology, Haifa 32000, Israel
{alexbron, bronstein}@ieee.org
[2] Department of Computer Science, Technion – Israel Institute of Technology, Haifa 32000, Israel
mzib@ee.technion.ac.il

Abstract. Presented here is the problem of recovering a dynamic image superimposed on a static background. Such a problem is ill-posed and may arise e.g. in imaging through semireflective media, in separation of an illumination image from a reflectance image, in imaging with diffraction phenomena, etc. In this work we study regularization of this problem in spirit of Total Variation and general sparsifying transformations.

1 Introduction

In this paper, we consider a problem of recovering images of two objects superimposed on each other, where one of the objects is static (background) and the other one is dynamic. Such problem can arise in imaging through semireflective media, in separation of an illumination image from a reflectance image, in imaging with varying diffraction phenomena, etc. An example of semireflective layers separation is shown in Figure 1. In such a setup, the camera observes a superposition of two layers: the real layer is the object (dog), seen through the glass. The virtual layer (girl) is formed by light reflected from the glass.

We assume that one of the objects is dynamic and that several time frames of the observed scene are available. Using this information, it is possible to recover both components by solving an underdetermined linear system. Since the problem is ill-posed, an appropriate regularization must be used.

Weiss [1] suggested a solution to this problem based on the observation that filtered version of an image is usually sparse, when the filter is a differentiation operator. However the suggested solution uses reconstruction in the domain of filtered images, which may suffer from noise amplification. Also explicit use of inverese filters restricts use of other (nonlinear) priors.

In our work we study more general regularizations of this problem, in spirit of Total Variation and general sparsifying transformations.

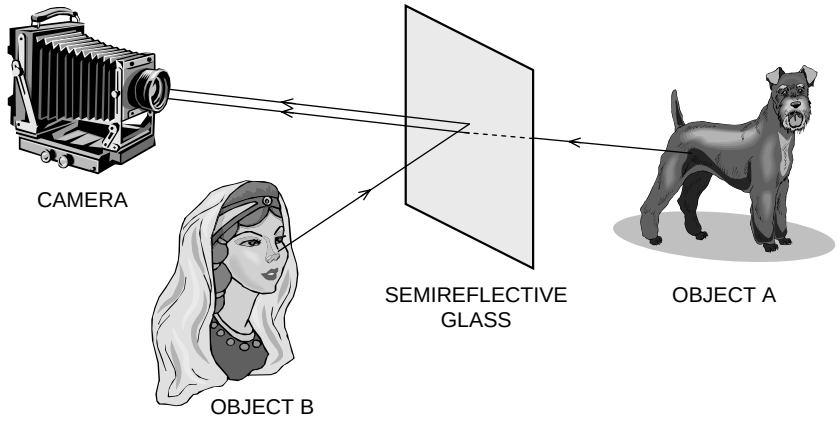

Fig. 1. Scheme of an optical setup involving a semireflector.

2 Regularized separation of layers

Let us be given a time sequence of T observations of the form

$$y^{(k)} = x^{(0)} + x^{(k)} + \xi^{(k)},\ k = 1, ..., T, \tag{1}$$

where $x^{(0)}$ is a $M \times N$ image of the static (background) object, $x^{(k)}$ are images of the dynamic object at different times, and $\xi^{(k)}$ is additive noise, which possibly contaminates the observations. An example of such a sequence is given in Figures 3–4. We will henceforth refer to the $T+1$ images $x^{(k)}, k = 0, ..., T$ as to *sources*. The problem of recovering $T+1$ unknown images from only T observed images is ill-posed. However, plausible separation results might be obtained if the solution is restricted to some class of images, to which the sources are believed *a priori* to belong. For simplicity of presentation, we assume that both the static and the dynamic sources obey the same prior (which is in general not necessary).

Assuming that the prior can be expressed via convex penalty function $\varphi(x)$, the separation problem can be formulated as finding such $x^{(k)}$'s that obey (1) up to some allowed discrepancy due to noise and slight deviations from the linear model, and minimize $\sum_k \varphi\left(x^{(k)}\right)$. This leads to the following constrained convex minimization problem:

$$\min_{x^{(0)},\ldots,x^{(T)}} \sum_{k=0}^{T} \varphi\left(x^{(k)}\right) \quad \text{s.t.} \quad \left\| x^{(0)} + x^{(k)} - y^{(k)} \right\|_2^2 \leq \beta$$
$$x^{(k)} \geq 0, \tag{2}$$

where β is usually chosen proportionally to the noise variance and the second constraint guarantees non-negativity of the estimated images. The latter can be reformulated as

unconstrained minimization of the convex function

$$f\left(x^{(0)},...,x^{(T)}\right) = \sum_{k=0}^{T}\varphi\left(x^{(k)}\right) + \lambda_1\sum_{k=1}^{T}\left\|x^{(0)}+x^{(k)}-y^{(k)}\right\|_2^2 + \lambda_2\sum_{i,j,k}\psi\left(x_{ij}^{(k)}\right), \tag{3}$$

where $\psi(t)$ is a penalty on negativity and λ_1, λ_2 are parameters.

2.1 Generalized total variation regularization

A powerful prior, suitable for wide classes of natural images is obtained when $\varphi(x)$ is chosen to be the *total variation* (TV) norm of the image x

$$\|x\|_{TV} = \sum_{i,j}\|\nabla x_{ij}\|_2 = \sum_{i,j}\sqrt{(\partial_x * x)_{ij}^2 + (\partial_y * x)_{ij}^2}, \tag{4}$$

where ∂_x and ∂_y are discrete derivative kernels in the x- and y-axis directions. TV has been successfully used for regularization in inverse problems, blind deconvolution, denoising, etc [2–5]. A more general form of the prior is the *generalized total variation* (GTV) [6], given by a general bivariate function of the form

$$\varphi(x) = \sum_{i,j} h\left((a*x)_{ij}, (b*x)_{ij}\right), \tag{5}$$

where a, b are some convolution kernels and e.g. $h(u,v) = \sqrt{u^2+v^2+\epsilon^2}$. In the latter case, the TV norm is a particular case obtained when $a = \partial_x$ and $b = \partial_y$.

2.2 Gradient and Hessian of $f\left(\boldsymbol{x^{(0)}},...,\boldsymbol{x^{(T)}}\right)$

Assuming the GTV prior (5), the gradient of $f\left(x^{(0)},...,x^{(T)}\right)$ from (3) is given by

$$\frac{\partial f}{\partial x_{ij}^{(m)}} = \sum_{k,l}\left(h_u\left(u_{kl}^{(m)}, v_{kl}^{(m)}\right)a_{k-i,l-j} + h_v\left(u_{kl}^{(m)}, v_{kl}^{(m)}\right)b_{k-i,l-j}\right) + 2\lambda_1\left(x_{ij}^{(0)} + x_{ij}^{(m)} - y_{ij}^{(m)}\right) + \lambda_2\psi'\left(x_{ij}^{(m)}\right) \tag{6}$$

for $m > 0$ and

$$\frac{\partial f}{\partial x_{ij}^{(0)}} = \sum_{k,l}\left(h_u\left(u_{kl}^{(0)}, v_{kl}^{(0)}\right)a_{k-i,l-j} + h_v\left(u_{kl}^{(0)}, v_{kl}^{(0)}\right)b_{k-i,l-j}\right) + 2\lambda_1\sum_{k=1}^{T}\left(x_{ij}^{(0)} + x_{ij}^{(k)} - y_{ij}^{(k)}\right) + \lambda_2\psi'\left(x_{ij}^{(0)}\right) \tag{7}$$

for $m = 0$, where $u^{(k)} = a * x^{(k)}$ and $v^{(k)} = b * x^{(k)}$. The first term accounting for the prior can be evaluated efficiently using FFT-based convolution. Since the kernels a

and b are usually significantly smaller compared to the source images $x^{(k)}$, the use of the overlap-and-add (OLA) method is especially advantageous.

The Hessian of $f\left(x^{(0)},...,x^{(T)}\right)$ is given by

$$\begin{aligned}\frac{\partial^2 f}{\partial x_{ij}^{(m)}\partial x_{i'j'}^{(m')}} = \sum_{k,l} &\Big(h_{uu}\left(u_{kl}^{(m)},v_{kl}^{(m)}\right)a_{k-i,l-j}a_{k-i',l-j'}\\ &+h_{vv}\left(u_{kl}^{(m)},v_{kl}^{(m)}\right)b_{k-i,l-j}b_{k-i',l-j'}\\ &+h_{uv}\left(u_{kl}^{(m)},v_{kl}^{(m)}\right)b_{k-i,l-j}a_{k-i',l-j'}\\ &+h_{uv}\left(u_{kl}^{(m)},v_{kl}^{(m)}\right)a_{k-i,l-j}b_{k-i',l-j'}\Big)\delta_{mm'}\\ &+2\lambda_1\gamma_{mm'}\delta_{ii'}\delta_{jj'}+\lambda_2\psi''\left(x_{ij}^{(m)}\right)\delta_{mm'}\delta_{ii'}\delta_{jj'},\end{aligned} \tag{8}$$

where $\delta_{mm'}$ is the Kroenecker delta and $\gamma_{mm'}=1$ for $m=m'$ or $m=0$ or $m'=0$ and 0 otherwise. In the following, for notation convenience we will denote the gradient by $\nabla_{x^{(0)},...,x^{(T)}}f(x^{(0)},...,x^{(T)})$ and the Hessian by $\nabla^2_{x^{(0)},...,x^{(T)}}f(x^{(0)},...,x^{(T)})$. Parsing the $T+1$ images to MN column vector, we can represent the gradient as a $(T+1)NM$ vector and the Hessian as a $(T+1)NM\times(T+1)NM$ matrix.

The Hessian has a $(T+1)\times(T+1)$ block structure with $MN\times MN$ blocks. The first term in (8) yields a band-diagonal structure of the diagonal blocks, where the number of the diagonals and the number of bands depend on the sizes of the kernels a and b. The second and the third terms account for a constant principal diagonal in the diagonal blocks of the Hessian, whereas the second term also accounts for a constant diagonal in the first row and column blocks. Typical Hessian structure is depicted in Figure 2. The sparse structure of the Hessian is very helpful for solving efficiently the Newton system, while carrying the optimization.

2.3 Sparsity-based priors

Another powerful class of priors on images is related to their sparse representation using some system of basis functions, or overcomplete "dictionaries", based on wavelets, curvelets, contourlets, etc. This paradigm is already used in image denoising and in solution of some inverse problems. Assume that the original images can be represented as

$$x^{(k)}=\sum_l c_l^{(k)}\phi_l$$

or in operator form

$$x^{(k)}=\Phi c^{(k)}$$

where the coefficients $c_l^{(k)}$ are sparse. The following regularized problem can be considered:

$$\min_c \sum_{k=0}^{T}\|c^{(k)}\|_1+\lambda_1\sum_{k=1}^{T}\left\|\Phi c^{(0)}+\Phi c^{(k)}-y^{(k)}\right\|_2^2+\lambda_2\sum_{i,j,k}\psi\left((\Phi c^{(k)})_{ij}\right),$$

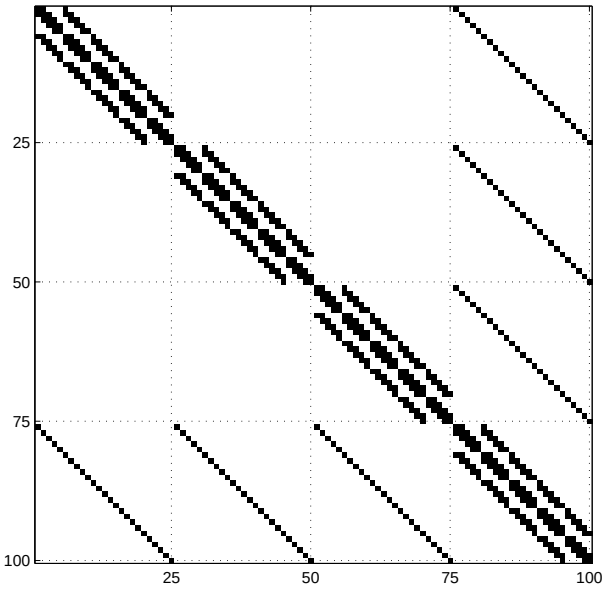

Fig. 2. Example of the Hessian sparse structure in a problem with 3 frames of size 5×5 and $a = \partial_x, b = \partial_y$. Black shows non-zero elements.

2.4 Minimization of $f\left(\boldsymbol{x^{(0)}, ..., x^{(T)}}\right)$

Since the function $f\left(x^{(0)}, ..., x^{(T)}\right)$ is convex with respect to $x^{(0)}, ..., x^{(T)}$, it can be minimized using standard convex optimization technique [7]. In our case, second-order Newton-type methods appear especially appealing. Let us denote by X the $(T+1)NM$ vector of variables $x^{(0)}, ..., x^{(T)}$ in column-stack representation. In the basic Newton method, the minimization of $f(X)$ is carried out by iteratively updating X in the following manner

$$X[k+1] = X[k] + \alpha[k]d[k] \tag{9}$$

where k denotes the iteration index, $\alpha(k)$ is the step size and $d(k)$ is the Newton direction, give by the solution of the *Newton system*:

$$\nabla^2 f(X[k])d[k] = -\nabla f(X[k]). \tag{10}$$

The Newton system, in turn, can be solved iteratively to some preset degree of accuracy, e.g. using the conjugate gradients method [7], which does not require an explicit Hessian computation, but rather computation of Hessian-vector products. Such computations are very efficient due to the sparse structure of the Hessian. This version of the Newton algorithm with approximate solution of the Newton system is often referred to as *inexact* or *truncated* Newton [8] method.

Since the function is convex, global convergence is guaranteed with any initialization. Yet, selecting an initialization which is sufficiently close to the solution, e.g. using the mixture images to initialize $x^{(0)}, ..., x^{(T)}$, faster convergence can be achieved.

3 Computational Results

In this work we present computational experiments with TV prior, leaving sparsity-based priors for future study. The proposed method was tested on synthetic data created by superposition of a static image of a female face and three frames showing a running dog (Figure 3). The observed result is a sequence of 3 frames (Figure 4).

The separation was carried out using the TV norm with smoothing parameter $\epsilon = 10^{-3}$ and the non-negativity penalty, with $\lambda_1 = 10^{-1}, \lambda_2 = 10^{-1}$. The mixture images $y^{(1)}, ..., y^{(3)}$ were used as the initialization for $x^{(1)}, ..., x^{(3)}$, and the image $y^{(1)}$ was used as the initialization for $x^{(0)}$. Optimization was carried out using the truncated Newton algorithm. The reconstructed images are shown in Figure 5.

Our algorithm provides a plausible separation of the background and the dynamic scene. Slight residuals of the dynamic scene are visible in the reconstructed image. These artifacts appear in regions where there is little or no motion in the dynamic scene, and thus the separation problem is ill-posed. The use of the TV prior results in a slight degradation of the texture details in the reconstructed dynamic scene.

4 Discussion

We presented an efficient solution of the ill-posed problem arising in separation of semireflective dynamic image from static background using TV prior. Further research should explore other types of priors, e.g. on coefficients of some decomposition (e.g. wavelet-type) of the images. Application to other optical problems should be considered as well. A potentially interesting application is separation of illumination and reflection components in pictures with multiple exposures [9].

References

1. Weiss, Y.: Deriving intrinsic images from image sequences. In: International Conf. Computer Vision (ICCV). (2001)
2. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D **60** (1992) 259–268
3. Blomgren, P., Chan, T.F., Mulet, P., Wong, C.: Total variation image restoration: numerical methods and extensons. In: Proc. IEEE ICIP. (1997)
4. Chan, T.F., Mulet, P.: Iterative methods for total variation image restoration. SIAM J. Num. Anal **36** (1999)
5. Chan, T.F., Wong, C.K.: Total variation blind deconvolution. In: Proc. ONR Workshop. (1996)
6. Bronstein, A.M., Bronstein, M.M., Zibulevsky, M., Zeevi, Y.Y.: Blind deconvolution of images using optimal sparse representations. IEEE Trans. Image Processing **14** (2005) 726–736
7. Bertsekas, D.P.: Nonlinear Programming (2nd edition). Athena Scientific (1999)
8. Nash, S.: A survey of truncated-Newton methods. Journal of Computational and Applied Mathematics **124** (2000) 45–59
9. Elad, M.: Retinex by two bilateral filters. In: Proc. Scale-Space. (2005)

Fig. 3. Source sequence. First column: static background image, second through fourth columns: three frames of the *dog* sequence.

Fig. 4. Observed mixtures sequence.

Fig. 5. Unmixed sequence.