# Separating Transparent Layers through Layer Information Exchange*

Bernard Sarel and Michal Irani

Dept. of Computer Science and Applied Mathematics,
Weizmann Institute of Science
Rehovot, ISRAEL
{bernard.sarel , michal.irani}@weizmann.ac.il

**Abstract.** In this paper we present an approach for separating two transparent layers in images and video sequences. Given two initial unknown physical mixtures, $I_1$ and $I_2$, of real scene layers, $L_1$ and $L_2$, we seek a layer separation which minimizes the structural correlations across the two layers, at *every* image point. Such a separation is achieved by transferring local grayscale structure from one image to the other wherever it is highly correlated with the underlying local grayscale structure in the other image, and vice versa. This bi-directional transfer operation, which we call the "layer information exchange", is performed on diminishing window sizes, from global image windows (i.e., the entire image), down to local image windows, thus detecting similar grayscale structures at varying scales across pixels. We show the applicability of this approach to various real-world scenarios, including image and video transparency separation. In particular, we show that this approach can be used for separating transparent layers in images obtained under different polarizations, as well as for separating complex *non-rigid* transparent motions in video sequences. These can be done without prior knowledge of the layer mixing model (simple additive, alpha-mated composition with an unknown alpha-map, or other), and under unknown complex temporal changes (e.g., unknown varying lighting conditions).

## 1 Introduction

The need to perform separation of visual scenes into their constituent layers arises in various real world applications (medical imaging, robot navigation, and others). This problem is challenging when the layers are transparent, thus generating complex superpositions of visual information. The problem is particularly challenging when the mixing process is an unknown, spatially varying, non-linear function, as is often the case in real-world transparent scenes.

A number of approaches to transparent layer separation have been proposed. Most of the approaches for separation of *still images* assume additive transparency with layer mixing functions which are uniform across the entire image (e.g., [8, 5, 7]). Spatially varying functions were handled by [3] assuming spareness of image derivatives . In the case of *video transparency* (where the transparent layers have different relative motions over time), the underlying assumption

* This research was supported in part by the Moross Laboratory at the Weizmann Institute of science.

is that dense correspondences can be pre-computed for each pixel in each layer across the entire sequence [9, 12]. These methods are therefore restricted to scenes with simple 2D parametric motions, which are easy to compute under transparency and provide dense correspondences. Non-parametric correspondences are handled in [10] assuming stereo images. None of the above methods can handle complex non-rigid motions. Szeliski *et al* [9, 10] further assume fixed mixing coefficients.

In this paper we address the problem of separation of two arbitrarily superimposed layers (either in images, or in video), without any prior knowledge about the mixing process. We assume that two different combinations of the layers (generated in an unknown fashion) are given to us, and use these to initiate the layer separation process. As will be shown later, two different combinations of layers are often available or otherwise easy to obtain in many real-world scenarios, making this approach practical.

Formally, and without loss of generality, we can phrase the problem as follows. Given two initial unknown physical mixtures, $I_1$ and $I_2$, of real scene layers, $L_1$ and $L_2$, produce approximations $\hat{L}_1$ and $\hat{L}_2$ such that some separation criterion is satisfied. The two mixtures $I_1$ and $I_2$, can be generally defined as,

$$\begin{aligned} I_1(i) &= \alpha_1(i) \cdot L_1(i) + \alpha_2(i) \cdot L_2(i) \\ I_2(i) &= \beta_1(i) \cdot L_1(i) + \beta_2(i) \cdot L_2(i) \end{aligned} \tag{1}$$

where the index $i$ denotes pixel position, and $\alpha_1(i)$,$\alpha_2(i)$,$\beta_1(i)$, and $\beta_2(i)$, are the unknown mixing functions (coefficients) which vary over pixel locations. In the simplest case, when the mixing is uniform and additive (as assumed in [9, 5, 8, 7]), the mixing functions reduce to constant coefficients; $\forall i \quad \alpha_1(i) \equiv \hat{\alpha}_1$, $\alpha_2(i) \equiv \hat{\alpha}_2$, $\beta_1(i) \equiv \hat{\beta}_1$ and $\beta_2(i) \equiv \hat{\beta}_2$. In natural scenes, however, such conditions are frequently violated. Smoothly varying glass opacity, window dirt, or images acquired through polarization filters, can produce varying mixing coefficients that vary over pixel locations. The formulation of Eq. (1) is general and captures a wide range of transparency models, including additive transparency with uniform mixing functions [9, 5, 8, 7], additive transparency with unknown alpha-matting (e.g., [12]), etc.

Having two initial combinations, $I_1$ and $I_2$, generated in an unknown fashion, we seek a layer separation into representations of $L_1$ and $L_2$ which minimizes the structural correlations across the two layers at *every* image point. Such a separation is achieved by transferring local structure from one image to the other wherever it is highly correlated with the underlying local structure in the other image, and vice versa. This bi-directional transfer operation, which we call the "layer information exchange", is performed on diminishing window sizes, from global image windows (i.e., the entire image) down to local image windows, thus detecting correlated structures at varying scales across pixel positions.

Two different initial combinations ($I_1$ and $I_2$) are available, e.g., when two images of the same transparent scene are taken with different polarizers (as in [5, 8]), or under different illuminations. However, our approach is not limited to those cases nor is it restricted to still imagery. When a single video camera records two transparent layers with different relative motions over time, and

when the motion of only *one* of those layers is computable (e.g., a 2D parametric motion), then such initial layer separation is possible. This can be done even if the second layer contains very complex non-rigid motions (e.g., running water). Moreover, the layer mixing process is not known and can possibly change over time, and other unknown complex temporal changes may also occur simultaneously (such as varying illumination and changing light reflections over time). Such examples are shown and discussed in the paper.

*This paper has three main contributions:* (i) The idea of "layer information exchange". (We also believe that this idea has applicability in disciplines of signal processing other than Computer Vision). (ii) To our best knowledge, this is the first time that video sequences containing *non-rigid* transparent motions have been separated (moreover, under unknown complex varying lighting conditions). (iii) Our approach provides a unified treatment to a wide range of transparency models, without requiring prior selection of the transparency model and the corresponding separation method. When the unknown mixing coefficients are spatially-invariant (i.e., only grayscale dependent, but independent of the pixel position), then our approach produces comparable results to Farid and Adelson's ICA-based separation [5]. However, when the mixing coefficients are spatially-varying (unknown) functions, our approach performs better. Similarly, if the motions of both transparent layers in a video sequence are easy to compute, then our approach compares to existing methods for separating video transparency [9, 12]. However, it performs better when one of the layers contains complex motions (such as non-rigid motions, 3D parralax) and other complex temporal changes.

The rest of the paper is organized as follows. In Section 2 we identify an information correlation measure which is best suited for the underlying problem. In Section 3 we introduce our layer information exchange process, which is used for recovering the separate layers. In Section 4 we show the applicability of the method to transparency separation in still images and in video sequences.

## 2 The Information Correlation Measure

There are various commonly used measures for correlating information across images. In this section we review some of their advantages and drawbacks, and identify a measure which is best suited for the task at hand.

The Mutual Information ($MI$) of two images ($f$ and $g$) captures the statistical correlation (or co-occurrence) of their grayscales: $MI(f,g) = H(f) + H(g) - H(f,g)$, where $H(f)$ is the entropy of the grayscale distribution in $f$, and $H(f,g)$ is the joint entropy [4]. Mutual Information can account for non-linear grayscale transformations which are *spatially invariant* (i.e. transformations which depend only on the grayscale value at a pixel, but not on the pixel position). However, it cannot account for *spatially varying* grayscale transformations which are pixel position dependent (such as the spatially varying mixing functions of Eq. (1)). In other words, if $\hat{f}$ is an image obtained from $f$ by some (non-linear) transformation on the histogram of $f$, then $MI(f,\hat{f}) = MI(f,f)$ (see Fig. 1.b and 1.c). However, if $\hat{f}$ is obtained from $f$ by some spatially varying (position-dependant)

(a) Original image | (b) Linear grayscale deformation | (c) Non-linear grayscale deformation | (d) Spatially varying (position dependent) deformation

$f_a$ | $f_b = (\frac{f_a}{2}) + 100$ | $f_c = f_a^3$ | $f_d(x,y) =$(see caption)

| Measure (normalized) | $\mathbf{f_a}$ **vs.** $\mathbf{f_a}$ | $\mathbf{f_a}$ **vs.** $\mathbf{f_b}$ | $\mathbf{f_a}$ **vs.** $\mathbf{f_c}$ | $\mathbf{f_a}$ **vs.** $\mathbf{f_d}$ |
|---|---|---|---|---|
| **MI** | 1.0000 | 1.0000 | 1.0000 | 0.3426 |
| **NGC** | 1.0000 | 1.0000 | 0.8329 | 0.8700 |
| **GNGC** | 1.0000 | 1.0000 | 0.9165 | 0.9981 |

(e)

**Fig. 1. Comparing different information correlation measures** *(a) Original image. (b) After a linear grayscale transformation. (c) After a nonlinear grayscale transformation. (d) After a spatially varying (i.e., position-dependent) grayscale transformation:* $f_d(x,y) = f_a \cdot (\sin(\frac{4\pi \cdot x}{n_x}) \sin(\frac{4\pi \cdot y}{n_y}) \cdot 0.333 + 0.667)$*, where* $n_x \times n_y$ *is the image size. (e) Comparing the information correlation between the original image* $f_a$ *and the transformed images* $(f_b, f_c, f_d)$ *under different measures (NGC, MI, GNGC – see Section 2). As can be seen, GNGC correlates extremely well across all transformations.*

grayscale transformation, then the mutual information of $f$ and $\hat{f}$ reduces significantly: $MI(f,\hat{f}) \ll MI(f,f)$, even though the geometric structures observed in $f$ and in $\hat{f}$ are highly correlated (see Fig. 1.d).

A different widely used information correlation measure is the Normalized Gray-scale Correlation ($NGC$): $NGC(f,g) = \frac{C(f,g)}{\sqrt{V(f) \cdot V(g)}}$, where $C(f,g) = \frac{1}{N}\sum_{j=1}^{N} f_j \cdot g_j - \bar{f} \cdot \bar{g}$ is the covariance of $f$ and $g$, $N$ is the number of pixels in $f$ ($f$ and $g$ are of the same size), $\bar{f}, \bar{g}$ are the average grayscale values of $f, g$, and $V(f) = \frac{1}{N}\sum_{j=1}^{N} f_j^2 - \bar{f}^2$ is the variance of $f$. $NGC$ can account only for *linear* grayscale transformations which are spatially invariant (i.e., only changes in the mean and variance of the intensity – see Fig. 1.b). Intuitively speaking, the *normalized correlation* (captured by $NGC$) can be regarded as a linear approximation of *statistical correlation* (captured by $MI$).

The above two measures require *global* grayscale correlations (whether normalized or statistical). We next define an information correlation measure which requires only local correlations, and can therefore account for a wide variety of grayscale variation (linear and non-linear), including *spatially-varying* (i.e., position-dependant) grayscale transformations. This measure, which we will refer to as the Generalized $NGC$ ($GNGC$) measure, is a weighted average of local

$NGC$ measures on small (typically 5x5) windows:

$$GNGC(f,g) = \frac{\sum_{i=1}^{N} NGC_i^2(f,g) \cdot (V_i(f) \cdot V_i(g))}{\sum_{i=1}^{N} (V_i(f) \cdot V_i(g))} = \frac{\sum_{i=1}^{N} C_i^2(f,g)}{\sum_{i=1}^{N} V_i(f) \cdot V_i(g)} \quad (2)$$

where $C_i(f,g)$ and $NGC_i(f,g)$ are, respectively, the local covariance and the local normalized correlation measure between two small corresponding windows $(5 \times 5)$ centered at pixels $i$ in images $f$ and $g$. In principle, one could define a similar global measure to that of Eq. (2) using a weighted sum of local $MI$ measures (instead of local $NGC$ measures). However, there is not enough grayscale statistics in small $5 \times 5$ windows, which is why we resort to the local $NGC$ measures. In case of color images, the sum is taken over all three color bands.

The normalized weighted sum in Eq. (2) takes into account the correlations of small corresponding windows across $f$ and $g$. These are weighted according to their reliability, which is measured by the grayscale variances in the local $(5 \times 5)$ windows. This captures correlations of small geometric features (under different grayscale transformations) without introducing numerical instabilities which are common to regular normalized correlation in small windows. Prominent geometrical features in the image are characterized by large local gray-scale variances and therefore contribute more to the global correlation ($GNGC$) measure, while flat gray-scale regions have small local grayscale variances, hence small weights.

Unlike the $MI$ measure, the $GNGC$ measure (Eq. (2)) captures also the statistical correlations between *geometric structures* in the image. It can therefore account for spatially varying non-linear grayscale transformations, such as the one showed in Fig. 1.d, whereas $MI$ cannot. The reason for this difference between the two measures, is that $MI$ requires *global* statistical correlation of grayscales across the two images (a condition which is violated under spatially-varying grayscale transformations), whereas $GNGC$ requires only *local* statistical correlation across the two images (but at every $5 \times 5$ window in the image). Similar measures to the $GNGC$ measure have been previously used for other tasks where correlation between geometric structures was needed (e.g., for multi-sensor alignment [6]), although in the past a regular integration of local correlation values for those tasks was typically used, whereas our global measure is a *weighted sum* of the local measures. This modification is crucial to the stability of the layer separation process.

Because $GNGC$ captures correlations of meaningful geometrical structures, it is therefore more suited for the problem at hand. Moreover, the $GNGC$ measure is easy to differentiate in order to derive an analytic solution to the layer separation problem, as will be shown in Section 3.

## 3 The Layer Information Exchange

Let $I_1$ and $I_2$ be two different combinations of two unknown layers $L_1$ and $L_2$, obtained in an unknown fashion (i.e., the coefficients $\alpha_1(i)$, $\alpha_2(i)$, $\beta_1(i)$ and $\beta_2(i)$ in Eq. (1) are unknown, spatially varying, non-linear mixing functions). We will obtain a separation of $I_1$ and $I_2$ into two layers $\hat{L}_1$ and $\hat{L}_2$ (which are visual representations of $L_1$ and $L_2$) by transferring information from $I_1$ to

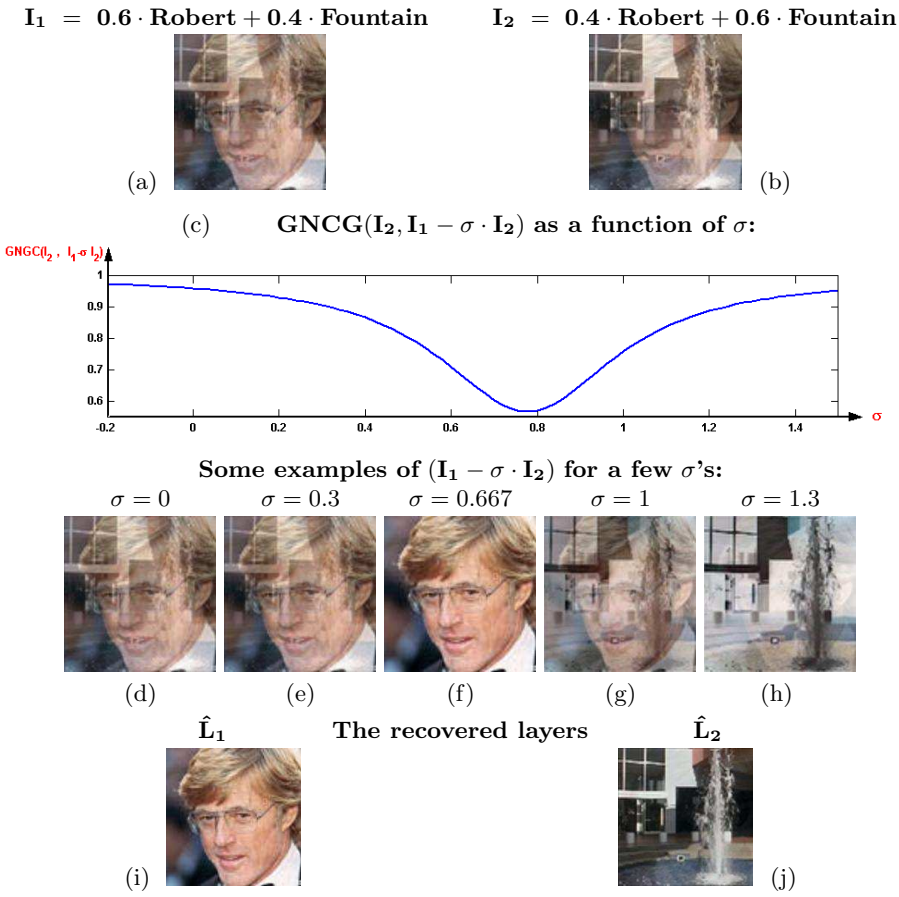


**Fig. 2. The Layer Information Exchange.** *(a)-(b) The initial mixtures $I_1$ and $I_2$. (c) Different values of $\sigma$ produce different degrees of information correlation between images $I_2$ and $I_1 - \sigma \cdot I_2$. (d)-(h) Examples of $I_1 - \sigma \cdot I_2$ for various values of $\sigma$. "Fountain" decreases until at $\sigma = 0.667$ it disappears completely, and when $\sigma$ is increased further, it becomes negative and the GNGC increases again. (i)-(j) The recovered layer separation using the algorithm described in Section 3.1.*

$I_2$, and vice versa, until the structural correlation between those two images is minimized. The information transfer is performed at different information scales, ranging from the entire image to small image windows. To explain this concept of "layer information exchange", let us first examine the simpler case of uniform mixing functions (i.e., constant unknown coefficients). We will later relax this assumption, and show how the process is generalized to spatially-varying non-linear mixing functions.

### 3.1 Handling Uniform Mixing Functions

Assuming uniform mixing functions, then Eq. (1) reduces to:

$$I_1(i) = \alpha_1 \cdot L_1(i) + \alpha_2 \cdot L_2(i) \quad , \quad I_2(i) = \beta_1 \cdot L_1(i) + \beta_2 \cdot L_2(i) \qquad (3)$$

There exists a constant scalar $\sigma$ such that $\hat{L}_1(i) = I_1(i) - \sigma I_2(i)$ will contain only the geometric structure of $L_1(i)$, without any trace of $L_2(i)$. For example, $\sigma = \frac{\alpha_2}{\beta_2}$ will lead to such a layer separation: $\hat{L}_1(i) = I_1(i) - \frac{\alpha_2}{\beta_2} I_2(i) = (\alpha_1 - \alpha_2 \frac{\beta_1}{\beta_2}) L_1(i)$. Namely, $L_1(i)$ is recovered up to a constant scale factor $(\alpha_1 - \alpha_2 \frac{\beta_1}{\beta_2})$. However, since $\alpha_1$, $\alpha_2$, $\beta_1$ and $\beta_2$ are not known, the transfer factor $\sigma$ is also unknown.

We do know, however, that for the correct transfer factor $\sigma$, the layer $L_2(i)$ will disappear in $\hat{L}_1(i)$, thus minimizing the structural correlation between $\hat{L}_1(i) = I_1(i) - \sigma I_2(i)$ and $I_2(i)$. This is visually shown in Fig. 2. We can therefore recover the transfer factor $\sigma$ (and accordingly the layer $L_1$, up to a scale), by minimizing the following objective function:

$$\sigma = argmin(GNGC(I_2, I_1 - \sigma I_2)) \tag{4}$$

Plugging in the definition of $GNGC$ from Eq. (2), results in an objective function which is quadratic in $\sigma$. Differentiating the above objective function with respect to $\sigma$ and equating to zero (i.e., $\frac{\partial}{\partial \sigma} GNGC(I_2, I_1 - \sigma I_2) = 0$), yields an analytic expression for $\sigma$:

$$\sigma = \frac{\sum_{i=1}^{N} C_i(I_1, I_2) \cdot V_i(I_2)}{\sum_{i=1}^{N} V_i^2(I_2)}, \tag{5}$$

where $C_i$ and $V_i$ are the local $(5 \times 5)$ covariances and variances as defined in Section 2. Having computed the transfer factor $\sigma$, we can recover the first layer (up to a scale):

$$\hat{L}_1 = I_1 - \sigma I_2,$$

and proceed to computing the second layer in the same way. The second layer

$$\hat{L}_2 = I_2 - \eta \hat{L}_1,$$

is recovered by seeking $\eta$ which minimizes $GNGC(\hat{L}_1, I_2 - \eta \hat{L}_1)$. In practice, we repeat this process a few times (typically 2 to 3 times), to obtain cleaner layer separation. At each iteration, the previously recovered $\hat{L}_1$ and $\hat{L}_2$ serve as the new mixtures. Namely, $\hat{L}_1^{k+1} = \hat{L}_1^k - \sigma^{k+1} \hat{L}_2^k$ , $\hat{L}_2^{k+1} = \hat{L}_2^k - \eta^{k+1} \hat{L}_1^{k+1}$ where $k$ is the iteration number.

We refer to the above procedure as the "layer information exchange", because at each step we transfer some portion of one image to the other. For example, the step $\hat{L}_1 = I_1 - \sigma I_2$ transfers some portion of $I_2$ to/from $I_1$ (depending on whether $\sigma$ is negative/positive). In the next step, a different portion of the new image $\hat{L}_1$ is transferred in the other direction, according to the magnitude and sign of $\eta$. Fig. 2.i and 2.j show the two layers recovered from images $I_1$ and $I_2$ (Fig. 2.a and 2.b) by applying the above information exchange procedure.

### 3.2 Generalizing to Spatially Varying Mixing Functions

So far we have assumed that the mixing coefficients $\alpha_1(i)$, $\alpha_2(i)$, $\beta_1(i)$ and $\beta_2(i)$ are constant. However, in most real-life scenarios, this is not true. To solve the separation problem for the case of spatially-varying mixing functions, we assume that if we use a small enough window $W_i$ around a pixel $i$, then within that region

of analysis the mixing coefficients are approximately uniform (although different from the mixing coefficients in other nearby pixels). In other words, the global layer exchange procedure described in Section 3.1 can be applied to a small local region of analysis $W_i$ to compute $\sigma(i)$ and $\eta(i)$ at the corresponding pixel $i$. These transfer factors are repeatedly computed for each pixel $i = 1..N$, using a window $W_i$ centered around each image pixel. This results in a *spatially-varying* layer information exchange: $\hat{L}_1(i) = I_1(i) - \sigma(i)I_2(i)$, and $\hat{L}_2(i) = I_2(i) - \eta(i)\hat{L}_1(i)$. This procedure is repeated iteratively: $\hat{L}_1^{k+1}(i) = \hat{L}_1^k(i) - \sigma^{k+1}(i)\hat{L}_2^k(i)$ , $\hat{L}_2^{k+1}(i) = \hat{L}_2^k(i) - \eta^{k+1}(i)\hat{L}_1^{k+1}(i)$ until $\sigma^k(i)$ and $\eta^k(i)$ are small enough (where $k$ is the iteration number).

Note that we are now dealing with two different types of local image windows: (i) the local region of analysis $W_i$ used of the piece-wise approximation of the mixing functions, and (ii) the small $5 \times 5$ window (mentioned in Section 2), which is used for obtaining local measurements (local $NGC$) to be summed for generating the global $GNGC$ measure. These $5 \times 5$ windows are the smallest reliable information elements over which the local $NGC$ measures are computed across the two images (regardless of whether the mixing functions are uniform or not). These local measures are then summed within the region of analysis, which is the entire image for the case of uniform mixing functions, and smaller $W_i$ in the case of spatially varying mixing functions.

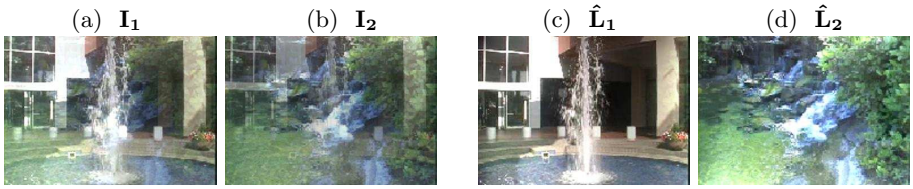


**Fig. 3. Handling spatially varying mixing functions.** *(a)-(b) The two mixtures $I_1$ and $I_2$ were obtained by mixing two images ("fountain" and "waterfall") with 4 different non-linear functions ($\alpha_1$ was a sinus, $\alpha_2$ and $\beta_1$ were two exponent functions, and $\beta_2$ was a constant function). (c)-(d) The recovered transparent layers using our global-to-local layer separation method described in Section 3.2 .*

Since we do not know ahead of time the degree of non-linearity of the mixing functions, the above local procedure is repeated using coarse-to-fine (i.e., large-to-small) regions of analysis $W_i$. We start the iterative process with $W_i$ being the entire image. This compensates for the case of globally uniform mixing functions (i.e., constant coefficients throughout the entire image). We then gradually decrease the window size $W_i$ to smaller and smaller windows (but not below $15 \times 15$, for numerical stability). This gradual process is aimed to assure that the resulting mixing functions remain as smooth as possible, whenever a smooth solution is a valid interpretation. Fluctuations from uniform/contant mixing functions occur when there is no simpler interpretation.

Fig. 3 shows an example of applying the above procedure to the pair of mixtures $I_1$ and $I_2$ (Figs. 3.a and 3.b). These images were generated with spatially varying non-linear mixing function/coefficients (see figure for more details). Fig. 3.c and 3.d show the resulting separation obtained using our layer

information exchange (without prior knowledge of the spatially varying mixing coefficients, of course). It has been able to completely separate the structures of the two layers.

## 4 Applications

The information exchange approach assumes that two different initial combinations ($I_1$ and $I_2$) of the unknown transparent layers ($L_1$ and $L_2$) are available, but the way in which $I_1$ and $I_2$ were generated from $L_1$ and $L_2$ is not known, and can be very complex. In this section we explore some cases where such initial combinations are readily available or else easy to extract, and show the applicability of our layer exchange approach for addressing these cases.


**Fig. 4. Recovering Transparent Layers from Polarized Images** *(a)-(b) Two real images obtained under different polarizations, showing the reflection of Sheila in a Renoir picture. (The images were taken from Farid [5].) (c)-(d) The recovered transparent layers using our layer separation method.*

### 4.1 Separating Layers in Polarized Images

Due to the physical nature of light polarization through reflecting and transmitting surfaces, two superimposed transparent layers differ in their polarization. Different mixtures ($I_1$ and $I_2$) of transparent scene layers can be obtained by changing the angle of a polarization filter in front of the camera (as in [5, 8]). Fig. 4 shows the result of applying our algorithm to a real pair of images of the same scene obtained with different polarizers. (These results are comparable to those of [5].)

### 4.2 Separating Non-Rigid Transparent Layers in Video

When a video camera records two transparent layers with different relative motions over time, and when the motion of *one* of those layers is easy to compute (e.g., if it is a 2D parametric motion), then such a layer separation is possible. This can be done even if the second layer contains very complex non-rigid motions (such as flickering fire, running water, walking people, etc.), the mixing process is not known and may be spatially varying (e.g., due to varying glass opacity or window dirt), and other temporal changes may occur simultaneously (such as varying illumination over time).

Such examples are shown in Fig. 5 (a simulated example) and in Fig. 7 (a real example). Fig. 5 shows a simulated example of an indoor scene with motion and

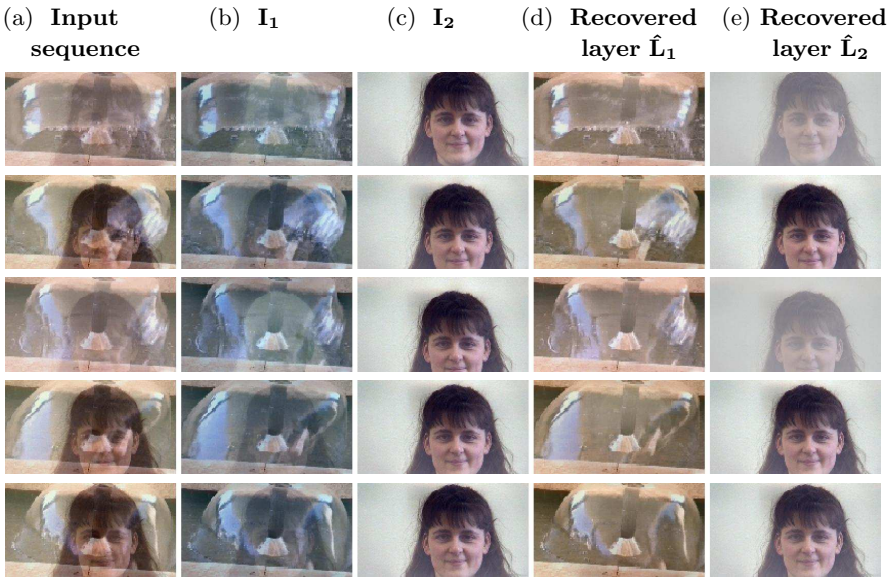


**Fig. 5. Separating non-rigid transparencies in video.** Column (a) *Five frames from the input movie (see text for details).* Columns (b)-(c) *The initial separation acquired by extracting the median image from the aligned sequence.* Columns (d)-(e) *The recovered layered. The residual traces of the woman which were visible in (b) are removed in (d), the true color of the fountain is recovered, and the temporal variations in the indoor illumination are recovered in (e). The video sequences can be viewed at* **http://www.wisdom.weizmann.ac.il/∼vision/TrasnparentLayers.html**.

varying illumination, reflected in a window through which a dynamic outdoor scene is visible. The input video sequence was generated by superimposing two video sequences: (i) an "indoor scene" video, showing a woman's head moving while the illumination changes over time (dimming and brightening of indoor illumination), reflected in a window, and (ii) an outdoor scene of a fountain displaying highly non-rigid complex motion, with changing specular reflections, etc. The left column of Fig. 5 displays some representative frames from the generated sequence. The woman's reflection is more visible when the illumination is darker, and is less visible when the illumination is brighter. The goal here was to separate this generated sequence into its original two layers (sequences, in this case): the outdoor scene (the fountain) with all its dynamics and specularities, and the indoor scene (the woman) with its motion and changing illumination.

In this case we have only one input (the video sequence of Fig. 5.a). To obtain *two* different initial layer mixtures ($I_1$ and $I_2$), we did the following: The woman's motion is a simple 2D parametric motion, which can be computed using one of the dominant motion estimation methods (e.g., [1, 2, 9]). This brings the woman into alignment. Now, using Weiss' method for extracting intrinsic images [11], we apply it to the aligned sequence. This process recovers a median image of the woman, and a residual image for each frame after removing the median image of the woman. These are displayed in the second and third columns of Fig. 5

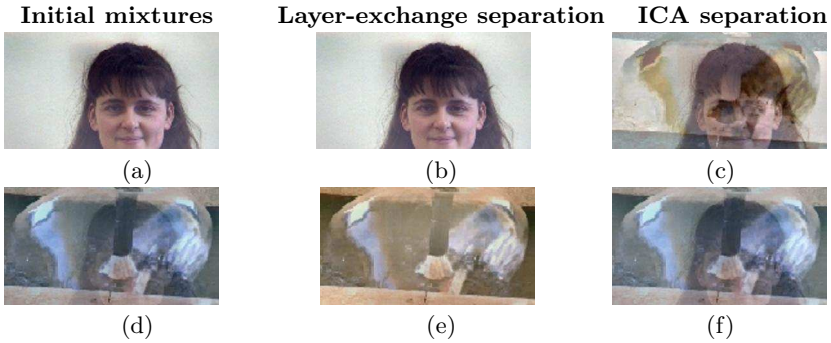


**Fig. 6. ICA vs. Information Exchange separation** *We compare results of applying the ICA-based separation [5, 13] to our layer-based separation displayed in Fig. 5. ICA was applied to the same initial mixture sequences (the "median" and the "residual" images in Figs. 5.b and 5.c). Almost all of the resulting frames displayed wrong separation. One such example is shown in the third column of this figure ((c) and (f)). For comparison, we display the corresponding frames of the initial mixture images (a) and (d), and our separation result (b) and (e).*

(after *un*warping the images to their original coordinate system according to the estimated 2D motion of the woman). Because the process of [11] results in a single intrinsic image, it does not capture any temporal changes. As a result, the woman's sequence in Fig. 5.c does not contain any of the changes in indoor illumination, and the "residual" sequence (Fig. 5.b) still contains a small residue of the woman (sometimes dark, sometimes bright), while the true colors of the fountain are lost.

Each pair of images in the second and third columns of Fig. 5 can be regarded as initial layer mixtures $I_1$ and $I_2$ (unknown and non-linear) for that time instance. These sequences ($I_1$ and $I_2$) are fed as the initial combinations to our layer exchange process. Results of the layer separation process are displayed in the last two columns of Fig. 5. Note that now the fountain sequence is fully recovered, with its true colors and no traces of the woman (Fig. 5.d), while the true changes in indoor illuminations have been recovered and automatically associated with the indoor woman sequence (Fig. 5.e).

The initial separation into a *"medians"* and *"residuals"* forms the initial mixtures $I_1$ and $I_2$ above. The (unknown) mixing functions which relate $I_1$ and $I_2$ to the original (unknown) layers (see Eq. (1)), *cannot* be assumed to be constant or position invariant. This is because the median operator is non-linear. Our Information Exchange approach handles this well (see Figs. 5.d and 5.e). However, the ICA-based separation [5, 13] does not perform well on these $I_1$ and $I_2$ , as can be seen in Figs. 6.c and 6.f. This is because it is not suited for the case of non-uniform spatially varying coefficients.

To our best knowledge, this is the first time videos containing *non-rigid* transparent motions have been separated (and moreover, under unknown varying lighting conditions). Current approaches for video transparency separation (e.g., [9, 10, 12]), assume that each layer moves rigidly, since dense correspondences of both layers across the sequence need to be recovered in those methods. We

(a) **Input sequence** (b) **Recovered Layer 1** (c) **Recovered Layer 2**

**Fig. 7. Separating non-rigid transparencies in video** *(a) Three frames from a real video sequence of the entrance hall of a building recorded through the building's swivelling glass door. The outdoor scene (including a running man and the camera tripod) are reflected from the swivelling door. The indoor scene includes a statue and a plant. (b) The first recovered layer (the outside scene). (c) The recovered interior hall with the statue. The video sequences can be viewed at* **http://www.wisdom.weizmann.ac.il/∼vision/TrasnparentLayers.html**.

currently need to compute only one of the motions, allowing the second motion to be arbitrarily complex.

Fig. 7 shows a real example of video transparency with non-rigid motions and changing effects of illumination. In this case, a still video camera recorded a scene with non-rigid human motions reflected in a swivelling glass door of an entry hall to a building. The reflected outdoor scene therefore appears moving, while the indoor scene is static. At the last part of the sequence, due to a strong reflections of light in the glass, the AGC (Automatic Gain Control) of the camera induced fluctuating changes in the dynamic range of the image. The left column of Fig. 7 displays three representative frames from the recorded sequence. As before, we used Weiss' method [11] for extracting the intrinsic image from the sequence. The median image was then removed from the sequence, producing a "residual" sequence. These were used as the initial combinations ($I_1$ and $I_2$) for

our layer exchange approach. The resulting separation into layers is displayed in the second and third columns of Figs. 7. The reflected scene was separated from the glass door, and the changing effects of illumination due to the change in aperture have also been recovered.

## 5 Conclusions

We presented an approach for separating two transparent layers through a process termed the "layer information exchange". Given two different (unknown complex) combinations of the layers, we recover the layers by gradually transferring information from one image to the other, until the structural correlation across the two images is minimized. The information transfer is done at different information scales, ranging from the entire image to small image windows.

We showed the applicability of this approach to various real-world scenarios, including image and video transparency separation. To our best knowledge, this is the first time that complex non-rigid transparent motions in video have been separated, without any prior knowledge of the layer mixing model, and under unknown complex temporal changes. We further showed that our approach to layer separation does equally well to ICA (Independent Component Analysis) when the mixing functions are spatially fixed (i.e., independent of the pixel position). However, when the mixing functions are more realistic spatially varying functions (i.e., vary as a function of pixel position), then our approach performs better than ICA. We believe that the applicability of this approach goes beyond analysis and separation of image layers, and can possibly be applied to separating other types of signals (such as acoustic signals, radar signals, etc.)

## References

1. J. R. Bergen, P. Anandan, K. J. Hanna, R. Hingorani: Hierarchical Model-Based Motion Estimation. ECCV 1992, pp. 237–252.
2. M. J. Black and P. Anandan: The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. CVIU, **63(1)**, 1996, pp. 75–104.
3. A. Bronstein, M. Bronstein, M. Zibulevsky and Y. Y. Zeevi: Separation of semireflective layers using Sparse ICA. Proc. ICASSP 2003, **3**, pp. 733–736.
4. T. Cover and J. Thomas: Elements of Information Theory, Wiley and Sons, 1991.
5. H. Farid and E.H. Adelson: Separating Reflections from Images by Use of Independent Components Analysis. JOSA, **16(9)** 1999, pp. 2136–2145.
6. M. Irani and P. Anandan: Robust Multi-Sensor Image Alignment. ICCV 1998, pp. 959–966.
7. A. Levin, A. Zomet, and Y. Weiss: Learning to perceive transparency from the statistics of natural scenes. NIPS 2002, pp.1247–1254.
8. Y. Y. Schechner, J. Shamir, and N. Kiryati: Polarization and statistical analysis of scenes containing a semi-reflector. JOSA A **17**, 2000, pp. 276–284.
9. R. Szeliski, S. Avidan, and P. Anandan: Layer Extraction from Multiple Images containing Reflections and Transparency. CVPR 2000, pp. 246–253.
10. Y. Tsin, S.B.Kang, and R. Szeliski: Stereo Matching with Reflections and Translucency. CVPR 2003, 702–709.
11. Y.Weiss: Deriving Intrinsic Images from Image Sequences. ICCV 2001, pp.68–75.
12. Y. Wexler, A. Fitzgibbon and A. Zisserman: Bayesian Estimation of Layers from Multiple Images. ECCV 2002, pp. 487–501.
13. "FastICA" package, downloaded from: http://www.cis.hut.fi/projects/ica/fastica/