Separating Reflections from Images Using Independent Components Analysis

Hany Farid and Edward H. Adelson

Perceptual Science Group, MIT, Cambridge, MA 02139

The image of an object can vary dramatically depending on lighting, specularities/reflections and shadows. It is often advantageous to separate these incidental variations from the intrinsic aspects of an image. Along these lines this paper describes a method for photographing objects behind glass and digitally removing the reflections off the glass leaving the image of the objects behind the glass intact. We describe the details of this method which employs simple optical techniques and independent components analysis (ICA) and show its efficacy with several examples.

1 Introduction

The image of an object can vary dramatically depending on lighting, specularities/reflections and shadows, and yet our recognition of objects is amazingly robust despite these incidental variations. Our visual system seems to separate the various components that contribute to the formation of an image, yielding stable and reliable percepts. One commonly occurring example is that of reflections from dielectric surfaces (e.g., glass) where our visual system seems to easily ignore reflections when, for example, viewing a painting framed behind glass. To facilite such tasks as object recognition, visual-based navigation, and scene segmentation we would like to design computer systems that have a similar ability to separate the incidental from the intrinsic aspects of an image. Along these lines, this paper describes a method for photographing objects viewed through glass and digitally removing the reflections off the glass leaving the image of the objects behind the glass intact. Although we will concentrate on this particular problem we believe that the general methods used may be useful for a broader class of image separation problems.

Light reflected off glass at an oblique angle is partially polarized suggesting that the strength of the reflection can be manipulated with a linear polarizer. The reflection can however only be completely eliminated when the viewing angle to the glass is at the Brewster's angle, typically a severe angle resulting in significant geometric distortions [1]. Consider the pair of images in Figure 1 of Renoir's *On the Terrace* framed behind glass with a reflection of a mannequin ("Sheila"). These images were photographed through a linear polarizer oriented to maximize the reflection (left) and minimize the reflection (right). The camera was oriented approximately 30° from frontal parallel. Note that even at the minimal orientation the reflection is quite salient.

There is some history in the computer vision community of trying to remove specular reflections from images. In these cases, researchers concerned themselves with non-planar reflective surfaces where the reflection is highly localized and poses a problem for various computer vision algorithms such as stereo and motion estimation. These approaches fall into one of several general categories: imposing a Lambertian assumption [2], color-based [3, 4, 5], polarization-based [6, 7] and combinations thereof [8]. Here we concern ourselves with the problem of removing reflections from a planar surface and take a completely different computational approach than previously suggested. Another important distinction is that in removing the reflections off the glass we leave the image of objects behind the glass intact.

Figure 1: Renoir's *On the Terrace* with a reflection of Sheila photographed through a linear polarizer at orthogonal orientations, maximizing and minimizing the reflection.

Let's first look more closely at the simple physics of reflections from planar surfaces. Shown in Figure 2 is an idealized example of a photograph of a painting behind a plate of glass. The final image is a linear combination of the light that is reflected by the painting and the light that is directly reflected by the glass. [1] The reflection from the glass will itself be an image of the scene in which the painting is viewed. The amount of light at a single point in the image can be expressed as:

$$y_1 \;=\; aP + bR, \tag{1}$$

where P and R are the amount of light contributed by the painting and reflection, and a and b are multiplicative constants. We would like to remove the contribution of the reflection R from the image y_1, but the above equation only provides a single constraint in four unknowns. Additional constraints may be added by exploiting the fact that reflections are partially polarized and that a linear polarizer can be used to adjust the relative strength of the reflection. With respect to Equation (1), the linear polarizer has the effect of manipulating the relative contributions of the painting and reflection. The second image takes the form:

$$y_2 \;=\; cP + dR. \tag{2}$$

The above equation provides another constraint, but two new unknowns have also been introduced, leaving us with a total of two constraints in six unknowns, and little hope of a solution without making further assumptions. It would of course be possible to manually

[1] In actuality, light is reflected off both the front and back face of the glass. In our conditions the pair of reflected images are nearly aligned so we assume the simplified model of an infinitely thin piece of glass with a single reflection.

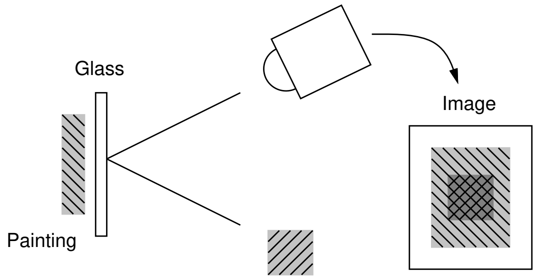

Figure 2: A photograph of a painting behind glass contains a superposition of the light that is reflected by the painting, and the light that is reflected directly off the glass.

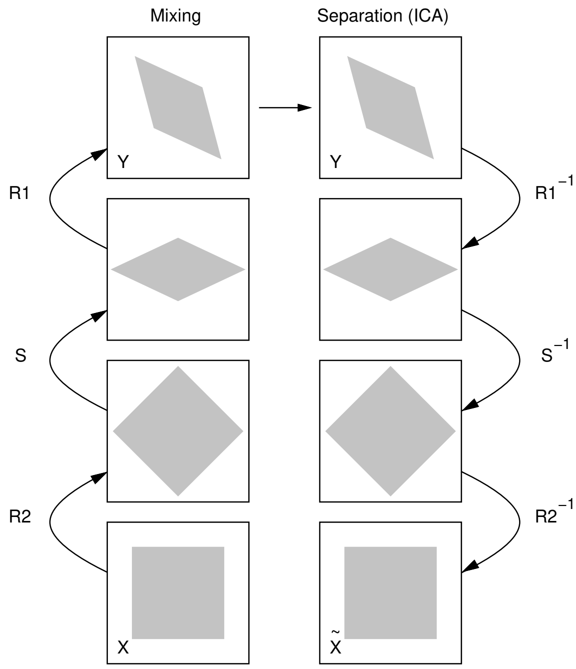

Figure 3: Shown at the bottom left is an idealized joint probability distribution for a pair of independent images. The linear mixing of these images, Equation (3), transforms this distribution, via a rotation, scaling, and rotation, from a square into a parallelogram (left column). The goal of ICA is to transform this parallelogram back into a square, thus yielding the original independent images (right column).

find the amount of image y_2 that needs to be subtracted from y_1 to remove the reflection R, but we are interested in an automatic procedure for accomplishing this.

In order to separate the reflection from the desired image we first photograph a scene through a linear polarizer at two distinct orientations. To solve the underconstrained set of equations we make the modest assumption that the image of the painting and reflection are independent. Intuitively this means that for each spatial position the pixel intensity in one image provides no predictive information about the pixel intensity in the second image. This is a reasonable assumption since there is no reason to expect a correlation between the image of objects behind the glass and the image of objects reflected by the glass. Given the linear model of image formation and this assumption of independence we perform independent components analysis (ICA) (e.g, [9, 10]) to separate the reflection from the desired image. In the next section the details of this statistical technique are outlined, and in the following section several examples of its efficacy are given.

2 Separating Images

The general problem of image separation can be stated as follows: given N distinct linear combinations of N images determine the original N images. For our application we can restrict ourselves to the case of just two images. Denoting these images in row vector form as x_1 and x_2, the linear mixing of these images can be expressed in matrix form as follows:

$$\begin{aligned} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} &= \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ Y &= MX \end{aligned} \tag{3}$$

where the matrix M embodies the linear mixing. Note that with this model it is assumed that the linear mixing is uniform over the entire image. The mixed images in Y each contain a linear combination of the source images in X. Our job is to recover the sources images from the mixed images. Of course given the full rank (i.e., invertible) matrix M it would be trivial to estimate the source images by left multiplying the mixed images with the inverse mixing matrix:

$$\tilde{X} = M^{-1}Y \tag{4}$$

But we don't typically know the mixing matrix, so our job will be to estimate it from the mixed images only.

Equation (3) provides two constraints in six unknowns and so cannot be solved without further assumptions. The first assumption we make is that the pair of source images are independent. Denoting X_1 and X_2 as the random variables from which the pixel intensities of source images x_1 and x_2 are drawn, this assumption can be expressed as $P(X_1, X_2) = P(X_1) \cdot P(X_2)$ (i.e., the joint probability distribution is separable). Although the constraint is expressed in terms of these continuous random variables we will typically work with the histograms of sampled images that, because of the dense sampling, are good approximations to the continuous probability distributions. The second modest assumption is that the mixing matrix M is full rank. With these two assumptions the general estimation problem is known as independent components analysis (ICA) of which there is a large and varied literature. Early contributors include [11, 9, 10, 12, 13, 14, 15, 16]. We present an analytic version of ICA based on higher-order statistical moments most similar to that of [9]. We provide a different formulation based on maximizing a series of error functions and give a simple and intuitive geometric interpretation of these steps.

We begin by attempting to gain some insight into the structure of the mixing matrix by decomposing it according to the singular value decomposition (SVD):

$$M = R_1 S R_2, \tag{5}$$

where R_1 and R_2 are orthonormal (rotation) matrices and S is a diagonal (scaling) matrix. Shown in Figure 3 is a geometric interpretation of the effects of each of these matrices on the joint probability distribution of the source images X. According to our assumption of independence this distribution should be separable; for illustration purposes consider the case when the marginal distributions are uniform, then the joint distribution is a square. The mixing matrix rotates, scales, and again rotates this distribution transforming the square into a parallelogram. Notice that the mixed images Y are no longer independent. The estimation of the independent source images reduces to determining how to transform the parallelogram back into a square. Or more generally, transforming the two-dimensional joint distribution into a separable product of one-dimensional distributions. As Figure 3 suggests this may be accomplished by applying the opposite rotations and scalings in reverse order.

The first step in separating the images is to apply a rotation that aligns the long and short axis of the parallelogram with the primary axis (Figure 3). These axes are easily determined since they are the axes with the maximal/minimal variance. Assuming zero mean measurements, the variance at an arbitrary orientation is given by:

$$E(\theta_1) = \sum_{i=1}^{N} \left[\begin{pmatrix} y_1(i) & y_2(i) \end{pmatrix} \begin{pmatrix} \cos(\theta_1) \\ \sin(\theta_1) \end{pmatrix} \right]^2 . \tag{6}$$

The axis of maximal variance is determined by finding the angle θ_1 that maximizes this error function (Figure 4). The axis of minimal variance will be orthogonal to this axis oriented at $\theta_1 - \pi/2$. These axes correspond to the principle axis as determined by principle components analysis (PCA). Since this error function is quadratic in its unknown it can be maximized analytically by differentiating with respect to θ_1, setting equal to zero and solving (Appendix A), to yield:

$$\theta_1 = \frac{1}{2} \tan^{-1} \left[\frac{\sum_{i=1}^{N} r^2(i) \sin(2\phi(i))}{\sum_{i=1}^{N} r^2(i) \cos(2\phi(i))} \right], \tag{7}$$

where, because it takes on a particularly simple form, the solution is given in polar coordinates $r(i) = y_1(i)^2 + y_2(i)^2$ and $\phi(i) = \tan^{-1}(y_2(i)/y_1(i))$. Then, the first rotation matrix in the separation is:

$$\tilde{R}_1 = \begin{pmatrix} \cos(\theta_1) & \sin(\theta_1) \\ -\sin(\theta_1) & \cos(\theta_1) \end{pmatrix}. \tag{8}$$

Following the first rotation, the now aligned parallelogram needs to be transformed into a diamond (Figure 3). More precisely, the axes need to be independently scaled so that the variance is rotationally invariant. The scaling of each axis is determined by first computing the variance along the axis of maximal and minimal variance, i.e., the axis oriented at θ_1 and $\theta_1 - \pi/2$:

$$s_1 = \sum_{i=1}^{N} \left[\begin{pmatrix} y_1(i) & y_2(i) \end{pmatrix} \begin{pmatrix} \cos(\theta_1) \\ \sin(\theta_1) \end{pmatrix} \right]^2 \tag{9}$$

$$s_2 = \sum_{i=1}^{N} \left[\begin{pmatrix} y_1(i) & y_2(i) \end{pmatrix} \begin{pmatrix} \cos(\theta_1 - \pi/2) \\ \sin(\theta_1 - \pi/2) \end{pmatrix} \right]^2, \tag{10}$$

and then the scaling matrix is constructed by placing the inverse variances along the diagonal:

$$\tilde{S} = \begin{pmatrix} s_1^{-1} & 0 \\ 0 & s_2^{-1} \end{pmatrix}. \tag{11}$$

Combined, the first rotation and scaling are equivalent PCA plus whitening. But notice from Figure 3 that this is insufficient for separating the mixed images into their independent components (the mixed images are only decorrelated, a necessary but insufficient condition). That is, in this example the joint probability distribution is in the shape of a diamond (i.e., is not separable), a final rotation is required to transform this diamond into a square, yielding the independent components (i.e., separability).

One approach to the determination of this final rotation is to find the orientation θ_2 that maximizes the

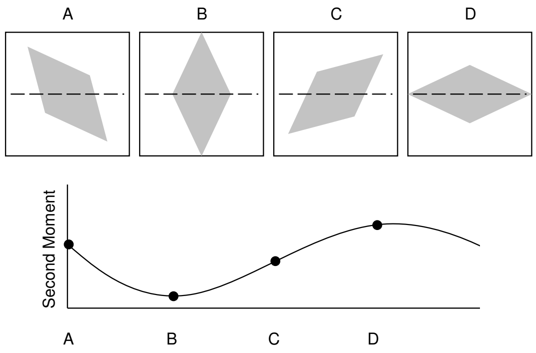

Figure 4: Shown is the variation in the second moment (variance) as the joint probability distribution is rotated and projected onto the horizontal axis (Equation (6)). The variance is minimal along the short axis of the parallelogram and maximal at the orthogonal orientation. Note that this error function is a 2-cycle sinusoid.

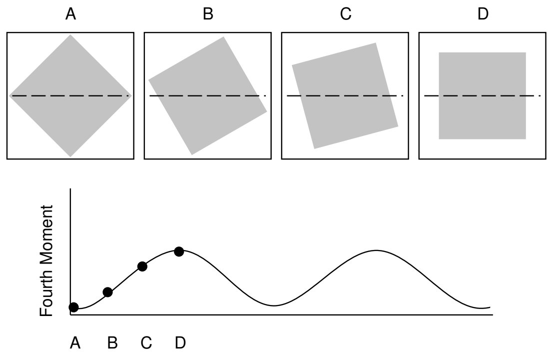

Figure 5: Shown is the variation in the fourth moment as the joint probability distribution is rotated and projected onto the horizontal axis (Equation (13)). The fourth moment is maximal at the orientation required to transform the diamond into a square, which in turn yields the independent images. Note that this error function is a 4-cycle sinusoid.

fourth statistical moment (Figure 5). The fourth moment at an arbitrary orientation is given by:

$$E(\theta_2) \;=\; \sum_{i=1}^{N} \left[\begin{pmatrix} y_1'(i) & y_2'(i) \end{pmatrix} \begin{pmatrix} \cos(\theta_2) \\ \sin(\theta_2) \end{pmatrix} \right]^4, \quad (12)$$

where y_1' and y_2' are the result of rotating and scaling the initial mixed images y_1 and y_2 according to Equations (8) and (11). Note that the above error function cannot be solved analytically. But, the following normalized fourth moment does lend itself to an analytic solution:

$$E'(\theta_2) \;=\; \frac{1}{||y_1'||^2 + ||y_2'||^2} E(\theta_2). \quad (13)$$

The axis where the fourth moment is maximal is determined by finding the angle θ_2 that maximizes this error function (Figure 5). As before, we differentiate with respect to θ_2, set equal to zero and solve (Appendix B) yielding the maximal solution:

$$\theta_2 \;=\; \frac{1}{4} \tan^{-1} \left[\frac{\sum_{i=1}^{N} r^2(i) \sin(4\phi(i))}{\sum_{i=1}^{N} r^2(i) \cos(4\phi(i))} \right], \quad (14)$$

again, for convenience, expressed in polar coordinates. The final rotation matrix then takes the form:

$$\tilde{R}_2 \;=\; \begin{pmatrix} \cos(\theta_2) & \sin(\theta_2) \\ -\sin(\theta_2) & \cos(\theta_2) \end{pmatrix}. \quad (15)$$

The estimation of the source images X from the mixed images Y is now a simple matter of applying the three matrices in Equations (8), (11), and (15):

$$\tilde{X} \;=\; (\tilde{R}_2 \tilde{S} \tilde{R}_1) Y \quad (16)$$

There are two inherent ambiguities in the recovery of the independent components. First is the ordering ambiguity, that is, the following mixings are indistinguishable:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \;=\; \begin{pmatrix} b & a \\ d & c \end{pmatrix} \begin{pmatrix} x_2 \\ x_1 \end{pmatrix} \quad (17)$$

Second is a scale ambiguity, that is the independent components can only be determined within a scale factor since, for example, the following mixings are also indistinguishable.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \;=\; \begin{pmatrix} a/\gamma & b/\delta \\ c/\gamma & d/\delta \end{pmatrix} \begin{pmatrix} \gamma x_1 \\ \delta x_2 \end{pmatrix} \quad (18)$$

For our purposes, the first of these ambiguities is not critical, and the second is dealt with by scaling the final images to fill the full intensity range.

3 Results

3.1 Synthetic

Our first experiment is intended to show the general efficacy of the image separation outlined in the previous section. Shown along the top row of Figure 6 are a pair of images and their normalized joint histogram (i.e., sampled joint probability distribution). In the next row are a pair of images formed by applying a random 2×2 mixing matrix as in Equation (3). In the subsequent

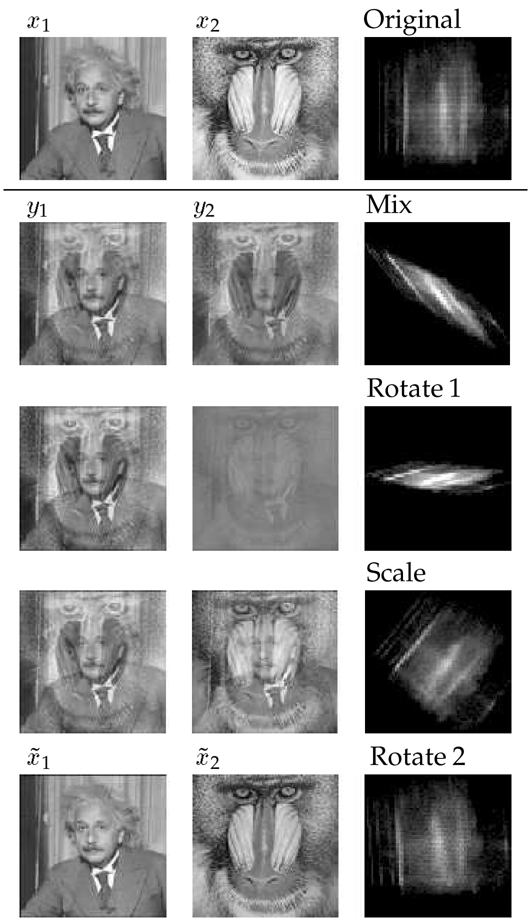

Figure 6: Along the top row are the original images of Einstein and Mandrill and their normalized joint histogram. The second row contains the mixed images and shown below are the intermediate steps leading to the separation of the mixed images. See also Table 1.

	Actual	Estimated
M	$\begin{pmatrix} 1.00 & -0.49 \\ 0.50 & -0.66 \end{pmatrix}$	$\begin{pmatrix} 1.00 & -0.63 \\ 0.49 & -0.79 \end{pmatrix}$
θ_1	35.7°	37.4°
s_1/s_2	4.41	4.55
θ_2	35.4°	41.4°

Table 1: Results from the separation of Einstein and Mandrill, Figure 6.

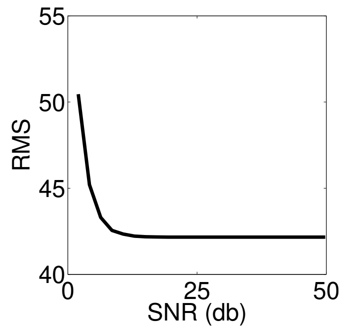

Figure 7: Separating Einstein and Mandrill in the presence of uniform zero-mean white noise. Shown is the root mean square (RMS) error between the separated and original images plotted as a function of the signal to noise ratio (SNR).

three rows are the results of applying the rotations and scalings specified by Equations (8), (11), (15) and (16) to this pair of images. As desired, the mixed images have been separated into their original independent components. Note also the similarity in the transformations of the normalized joint histogram with the idealization of Figure 3. Shown in Table 1 are the initial and estimated parameters of the mixing matrix. The pair of rotations are determined correctly within 1.7°and 6°, and the scaling within 3%.

To measure the sensitivity to noise, varying amounts of uniform zero-mean noise were added to the mixed images y_1 and y_2 before separating the independent components. Shown in Figure 7 is the root mean square (RMS) error between the separated and original images. The RMS errors are averaged over one hundred independent trials.

3.2 Natural

In the next set of experiments we photographed a painting framed behind glass with the reflection of a mannequin ("Sheila"), Figure 8. The painting with reflection was photographed twice through a linear polarizer oriented so that the reflection was maximized, and then at the orthogonal direction to minimize the reflection. We chose the maximum and minimum orientations to reinforce the fact that a linear polarizer alone is insufficient for removing the reflection. In general, any two distinct orientations could be used that are not symmetric about the polarization angle of the reflected light. We used a color digital video camera (Canon Optura DV, Canon Inc.) whose video signal was digitized through a S-video connection onto a Silicon Graphics computer. The camera was calibrated to ensure a linear response.

Figure 8: Renoir's *On the Terrace*, Sheila and Sheila's reflection.

Each scene was photographed for one-third of a second (10 frames) these frames were averaged in order to reduce sensor noise. The final three channel (RGB) color images have a spatial resolution of 640×480 pixels. In our results the entire three channel image was used in computing the independent sources.

Results are shown in Figures 9 and 10. Shown along the top row of each of these figures are the initial images and shown along the bottom row are the results of separating the independent components. In between are the intermediate steps leading to the separation as specified by Equations (8), (11), (15) and (16). Note how the normalized joint histogram (third column) transforms from a parallelogram to a square similar to the idealization of Figure 3, thus yielding the independent components - the painting and the reflection of Sheila. In the last experiment the reflections in the photograph of a storefront window are removed, shown in Figure 11 are the initial and separated images.

4 Discussion

We have developed a simple and effective method for separating reflections from images. This technique begins with a pair of images taken through a linear polarizer at two distinct orientations. The reflection is separated from the image by applying an analytic version of independent components analysis (ICA) based on higher-order statistical moments. A real-time implementation may be realized by synchronizing the image capture of alternate or even/odd frames with a liquid crystal polarizer (e.g, [17]).

There are of course several natural extensions of our work that will undoubtedly generalize its applicability. Most notably, the linear mixing model of Equa-

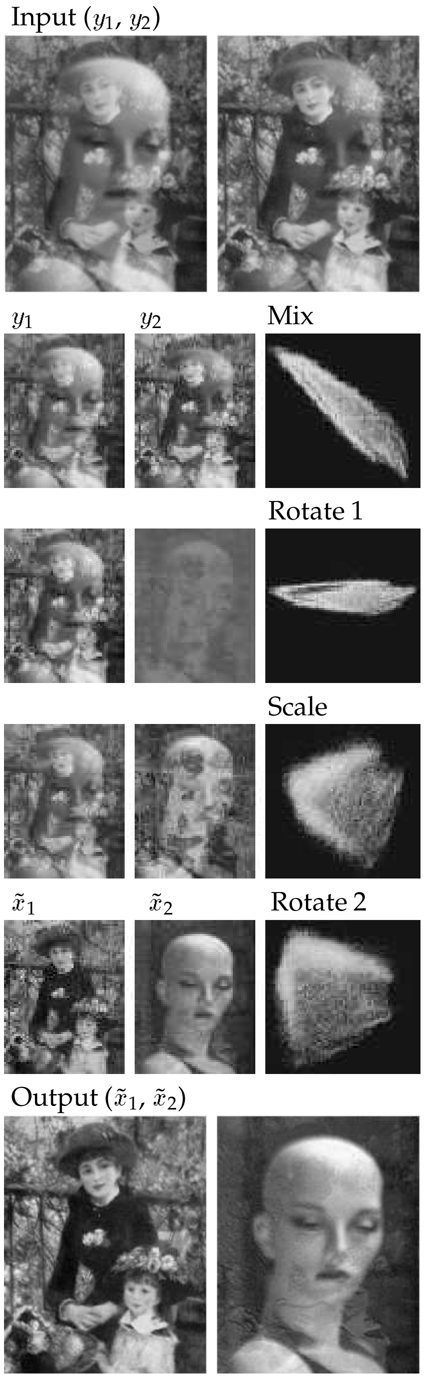

Figure 9: Along the top row are a pair of images of Renoir's *On the Terrace* with a reflection of Sheila photographed through a linear polarizer at orthogonal orientations. Along the bottom row are the independent components. Also shown are the intermediate steps leading to the separation of the independent components. The third column shows the normalized joint histogram of the pair of images to its left.

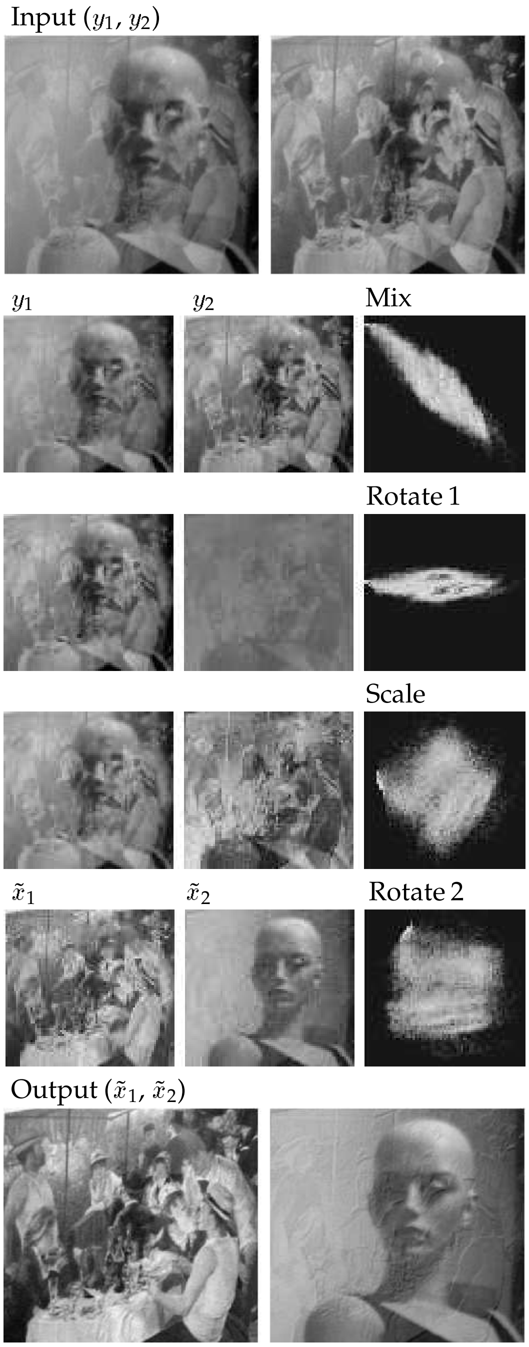

Figure 10: Along the top row are a pair of images of Renoir's *Lunching on the Boating Party* with a reflection of Sheila photographed through a linear polarizer at orthogonal orientations. Along the bottom row are the independent components. Also shown are the intermediate steps leading to the separation of the independent components. The third column shows the normalized joint histogram of the pair of images to its left.

Figure 11: Shown along the top row are the pair of images photographed through a linear polarizer at orthogonal orientations, and shown along the bottom row are the separated components.

tion (3) assumes a spatially uniform linear combination of source images that is unlikely to be true. The mixing will vary spatially because the polarization of reflected light depends on the angle to the camera, and this angle varies across the field of view. To account for this variation, our basic approach could be used in conjunction with a mixture model approach to fit multiple linear models [18]. Another possible extension would be to employ a more generic version of ICA that allows for a redundant system with more measurements than unknowns (e.g., [19]). In our case, this would mean photographing a scene at multiple polarization angles. Note though that since the polarization of a scene is fully characterized by three distinct orientations, there will be a point of diminishing returns. And lastly, our separation method could be applied selectively, for example when photographing a scene that is not completely behind glass. In this case, the pair of images can be subtracted, and only regions with significant differences will be considered to have a reflected component.

This technique may also be useful in other domains.

For example, reflections from the windshield of an autonomous vehicle with inboard cameras may cause problems for visually guided navigation systems. Another possible application is in the field of surveillance where activities behind a reflective window may be revealed.

Acknowledgments

The authors gratefully acknowledge support from NIH Grant EY11005-04 and MURI Grant N00014-95-1-0699. We would also like to thank Mary Bravo and Eero Simoncelli for insightful discussions and comments.

References

[1] M. Born and E. Wolf. *Principles of Optics*. Pergamon, London, 1965.

[2] G. Brelstaff and A. Blake. Detecting specular reflections using lambertian constraints. In *International Conference on Computer Vision*, pages 297–302, Tampa, FL, December 1988.

[3] S. Shafer. Using color to separate reflection components. *Color Research and Applications*, 10:210–218, 1985.

[4] R. Bajcsy, S.W. Lee, and A. Leonardis. Color image segmentation with detection of highlights and local illumination induced by interreflections. In *Proceedings of International Conference on Pattern Recognition*, pages 785–790, Atlantic City, NJ, June 1990.

[5] G.J. Klinker, S.A. Shafer, and T. Kanade. The measurement of highlights in color images. *International Journal of Computer Visions*, 2(1):7–32, 1990.

[6] L.B. Wolff. Polarization-based material classification from specular reflection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(11):1059–1071, 1990.

[7] L.B. Wolff and T.E. Boult. Constraining object features using a polarization reflectance model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(7):635–657, 1991.

[8] S.K. Nayar, X. Fang, and B. Terrance. Separation of reflection components using color and polarization. *International Journal of Computer Visions*, 21(3):163–186, 1997.

[9] J. Cardoso. Source separation using higher order moments. In *International Conference on Acoustics, Speech and Signal Processing*, pages 2109–2112, 1989.

[10] P. Comon. Separation of stochastic processes. In *Workshop on Higher-Order Spectral Analysis*, pages 174–179, Vail, Colorado, June 1989.

[11] Y. Bar-Ness. Bootstrapping adaptive interference cancelers: Some practical limitations. In *The Globecom Conference*, pages 1251–1255, 1982.

[12] Y. Inouye and T. Matsui. Cumulant based parameter estimation of linear systems. In *Workshop on Higher-Order Spectral Analysis*, pages 180–185, Vail, Colorado, June 1989.

[13] M. Gaeta and J.L. Lacoume. Source separation without a priori knowledge: The maximum likelihood solution. In Masgrau Torres and Lagunas, editors, *Proceedings EUSIPCO Conference*, pages 621–624, Barcelona, 1990.

[14] V.C. Soon, L. Tong, Y.F Huang, and R. Liu. An extended fourth order blind identification algorithm in spatially correlated noise. In *International Conference on Acoustics, Speech and Signal Processing*, pages 1365–1368, 1990.

[15] C. Jutten and J. Herault. Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *IEEE Transactions on Signal Processing*, 24(1):1–10, 1991.

[16] J.L. Lacoume and P. Ruiz. Separation of independent sources from correlated inputs. *IEEE Transactions on Signal Processing*, 40(12):3074–3078, 1992.

[17] L.B. Wolff, T.A. Mancini, P. Pouliquen, and A.G. Andreou. Liquid crystal polarization camera. *IEEE Transactions on Robotics and Automation*, 13(2):195–203, 1997.

[18] R.A. Redner and H.F. Walker. Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review*, 26:195–239, 1994.

[19] J. Cardoso. Super-symmetric decomposition of the fourth-order cumulant tensor. Blind identification of more sources than sensors. In *International Conference on Acoustics, Speech and Signal Processing*, pages 3109–3112, 1991.

Appendix A

The second statistical moment (variance) at an arbitrary angle θ_1 is given by:

$$
\begin{aligned}
E(\theta_1) &= \sum_{i=1}^{N} \left[\begin{pmatrix} y_1(i) & y_2(i) \end{pmatrix} \begin{pmatrix} \cos(\theta_1) \\ \sin(\theta_1) \end{pmatrix} \right]^2 \\
&= \sum_{i=1}^{N} \left[y_1(i)\cos(\theta_1) + y_2(i)\sin(\theta_1) \right]^2 \\
&= \sum_{i=1}^{N} y_1^2(i)\cos^2(\theta_1) + 2y_1(i)y_2(i)\cos(\theta_1)\sin(\theta_1) + y_2^2(i)\sin^2(\theta_1). \qquad (19)
\end{aligned}
$$

The angle that maximizes the variance can be determined by first differentiating this error function:

$$
\begin{aligned}
\frac{dE(\theta_1)}{d\theta_1} &= \sum_{i=1}^{N} -2y_1^2(i)\sin(\theta_1)\cos(\theta_1) + 2y_1(i)y_2(i)(\cos^2(\theta_1) - \sin^2(\theta_1)) + 2y_2(i)^2\sin(\theta_1)\cos(\theta_1) \\
&= 2\sum_{i=1}^{N} (y_2^2(i) - y_1^2(i))\sin(\theta_1)\cos(\theta_1) + y_1(i)y_2(i)(\cos^2(\theta_1) - \sin^2(\theta_1)) \\
&= 2\sum_{i=1}^{N} (y_2^2(i) - y_1^2(i))(\tfrac{1}{2}\sin(2\theta_1)) + y_1(i)y_2(i)(\tfrac{1}{2}(1+\cos(2\theta_1)) - \tfrac{1}{2}(1-\cos(2\theta_1))) \\
&= \sum_{i=1}^{N} (y_2^2(i) - y_1^2(i))\sin(2\theta_1) + 2y_1(i)y_2(i)\cos(2\theta_1), \qquad (20)
\end{aligned}
$$

setting equal to zero and solving:

$$
\begin{aligned}
\frac{dE(\theta_1)}{d\theta_1} &= 0 \\
\frac{\sin(2\theta_1)}{\cos(2\theta_1)} &= \frac{-2\sum_{i=1}^{N} y_1(i)y_2(i)}{\sum_{i=1}^{N} y_2^2(i) - y_1^2(i)} \\
\theta_1 &= \frac{1}{2}\tan^{-1}\left[\frac{-2\sum_{i=1}^{N} y_1(i)y_2(i)}{\sum_{i=1}^{N} y_2^2(i) - y_1^2(i)} \right]. \qquad (21)
\end{aligned}
$$

And finally converting into polar coordinates with $y_1(i) = r(i)\cos(\phi(i))$ and $y_2(i) = r(i)\sin(\phi(i))$ yields the maximal solution as in Equation (7):

$$
\begin{aligned}
\theta_1 &= \tan^{-1}\left[\frac{-2\sum_{i=1}^{N} r(i)\cos(\phi(i))r(i)\sin(\phi(i))}{\sum_{i=1}^{N} r^2(i)\sin^2(\phi(i)) - r^2(i)\cos^2(\phi(i))} \right] \\
&= \tan^{-1}\left[\frac{-2\sum_{i=1}^{N} \frac{1}{2}r^2(i)\sin(2\phi(i))}{\sum_{i=1}^{N} r^2(i)(\frac{1}{2}(1-\cos(2\phi(i))) - \frac{1}{2}(1+\cos(2\phi(i))))} \right] \\
&= \frac{1}{2}\tan^{-1}\left[\frac{\sum_{i=1}^{N} r^2(i)\sin(2\phi(i))}{\sum_{i=1}^{N} r^2(i)\cos(2\phi(i))} \right]. \qquad (22)
\end{aligned}
$$

Appendix B

The fourth statistical moment at an arbitrary angle θ_2 is given by:

$$
\begin{aligned}
E(\theta_2) &= \sum_{i=1}^{N} \left[\begin{pmatrix} y_1(i) & y_2(i) \end{pmatrix} \begin{pmatrix} \cos(\theta_2) \\ \sin(\theta_2) \end{pmatrix} \right]^4 \\
&= \sum_{i=1}^{N} \left[y_1(i)\cos(\theta_2) + y_2(i)\sin(\theta_2) \right]^4 \\
&= \sum_{i=1}^{N} y_1^4(i)\cos^4(\theta_2) + 4y_1^3(i)y_2(i)\cos^3(\theta_2)\sin(\theta_2) + 6y_1^2(i)y_2^2(i)\cos^2(\theta_2)\sin^2(\theta_2) \\
&\quad + 4y_1(i)y_2^3(i)\cos(\theta_2)\sin^3(\theta_2) + y_2^4(i)\sin^4(\theta_2) \\
&= \sum_{i=1}^{N} \tfrac{1}{8}y_1^4(i)(3 + 4\cos(2\theta_2) + \cos(4\theta_2)) + y_1^3(i)y_2(i)(\sin(2\theta_2) + \tfrac{1}{2}\sin(4\theta_2)) + \tfrac{6}{4}y_1^2(i)y_2^2(i)(\tfrac{1}{2} - \tfrac{1}{2}\cos(4\theta_2)) \\
&\quad + y_1(i)y_2^3(i)(\sin(2\theta_2) - \tfrac{1}{2}\sin(4\theta_2)) + \tfrac{1}{8}y_1^2(i)(3 - 4\cos(2\theta_2) + \cos(4\theta_2)). \qquad (23)
\end{aligned}
$$

The angle that maximizes the fourth moment can be determined by differentiating this error function:

$$
\begin{aligned}
\frac{dE(\theta_2)}{d\theta_2} &= \sum_{i=1}^{N} \tfrac{1}{8}y_1^4(i)(-8\sin(2\theta_2) - 4\sin(4\theta_2)) + y_1^3(i)y_2(i)(2\cos(2\theta_2) + 2\cos(4\theta_2)) + \tfrac{3}{2}y_1^2(i)y_2^2(i)(2\sin(4\theta_2)) \\
&\quad + y_1(i)y_2^3(i)(2\cos(2\theta_2) - 2\cos(4\theta_2)) + \tfrac{1}{8}y_1^2(i)(8\sin(2\theta_2) - 4\sin(4\theta_2)) \qquad (24)
\end{aligned}
$$

Note that this error function cannot be solved analytically, but that the following normalized error function does lend itself to an analytic solution:

$$
E'(\theta_2) = \frac{1}{||y_1||^2 + ||y_2||^2} E(\theta_2) \qquad (25)
$$

and

$$
\begin{aligned}
\frac{dE'(\theta_2)}{d\theta_2} &= \frac{1}{||y_1||^2 + ||y_2||^2} \frac{dE(\theta_2)}{d\theta_2} \\
&= \sum_{i=1}^{N} \frac{1}{y_1(i)^2 + y_2(i)^2} \left[(y_2^4(i) - y_1^4(i))\sin(2\theta_2) + (2y_1(i)y_2^3(i) + 2y_1^3(i)y_2(i))\cos(2\theta_2) \right] \\
&\quad + \frac{1}{y_1(i)^2 + y_2(i)^2} \left[(-\tfrac{1}{2}y_1^4(i) - \tfrac{1}{2}y_2^4(i) + 3y_1^2(i)y_2^2(i))\sin(4\theta_2) + (2y_1^3(i)y_2(i) - 2y_1(i)y_2^3(i))\cos(4\theta_2) \right] \\
&= \sum_{i=1}^{N} \frac{1}{y_1(i)^2 + y_2(i)^2} \left[(-\tfrac{1}{2}y_1^4(i) - \tfrac{1}{2}y_2^4(i) + 3y_1^2(i)y_2^2(i))\sin(4\theta_2) + (2y_1^3(i)y_2(i) - 2y_1(i)y_2^3(i))\cos(4\theta_2) \right] \qquad (26)
\end{aligned}
$$

where the $2\theta_2$ terms cancel from the earlier maximization (see Appendix A). This error function can now be maximized by setting equal to zero and solving:

$$
\begin{aligned}
\frac{dE'(\theta_2)}{d\theta_2} &= 0 \\
\frac{\sin(4\theta_2)}{\cos(4\theta_2)} &= \frac{-\sum_{i=1}^{N} \frac{1}{y_1(i)^2 + y_2(i)^2}(2y_1^3(i)y_2(i) - 2y_1(i)y_2^3(i))}{\sum_{i=1}^{N} \frac{1}{y_1(i)^2 + y_2(i)^2}(-\frac{1}{2}y_1^4(i) - \frac{1}{2}y_2^4(i) + 3y_1^2(i)y_2^2(i))} \\
\theta_2 &= \frac{1}{4}\tan^{-1}\left[\frac{-\sum_{i=1}^{N} \frac{1}{y_1(i)^2 + y_2(i)^2}(2y_1^3(i)y_2(i) - 2y_1(i)y_2^3(i))}{\sum_{i=1}^{N} \frac{1}{y_1(i)^2 + y_2(i)^2}(-\frac{1}{2}y_1^4(i) - \frac{1}{2}y_2^4(i) + 3y_1^2(i)y_2^2(i))} \right]. \qquad (27)
\end{aligned}
$$

And converting into polar coordinates with $y_1(i) = r(i)\cos(\phi(i))$ and $y_2(i) = r(i)\sin(\phi(i))$ yields the maximal solution as in Equation (14):

$$\theta_2 = \frac{1}{4}\tan^{-1}\left[\frac{\sum_{i=1}^{N} r^2(i)\sin(4\phi(i))}{\sum_{i=1}^{N} r^2(i)\cos(4\phi(i))}\right]. \tag{28}$$