Color Balance in Conventional Halftoning

K. Daels, P. Delabastita *

Keywords: color balance, halftones, moiré, register, rosette.

Abstract. The rosette structure in the traditional halftoning technique for four color printing is normally fixed, but it can also be periodically variant when moiré is present. The two extreme appearances of the rosette have been described as "clear" and "dot" centered. In this paper the morphology of the rosette is discussed in depth, and its relation is investigated with color balance. An improvement is presented of the traditional halftoning system that both minimizes the visibility of the rosette and optimizes color balance.

Introduction

The goal of rotating the halftone screens in traditional four color printing is to provide a pseudo randomization of the relative position of the halftone dots in the different separations in order to make the average amount of overlap between the dots - and hence the color balance - less registration dependent.

The problem of moiré in four color printing was studied in ref. (1). It was shown by means of a vector diagram in the frequency domain that no low frequency moiré occurs if the screening angles and frequencies of the cyan, magenta and black separations are selected so that their corresponding vectors form a closed triangle. (The light absorption of the fourth color, yellow, is usually low enough so that its interferences with the other colors are not objectionable, so it is left out from this discussion.) This condition is particularly met in the traditional screening system in which the frequencies of these separations are all equal, and in which the angles are different by exactly 120 degrees. If for some reason (for example angular misregistration) the vectors in the frequency diagram do not form a closed triangle, a low frequency moiré will occur of which the angle and frequency are predicted by the opening in the triangle.

* Agfa-Gevaert N.V., Septestraat 27, B-2640 MORTSEL (Belgium).
uucp: k_daels@roam.agfa.be, p_delabastita@roam.agfa.be

The analysis in ref. (1) did not include the effects of the harmonics of the screens, and neither did it extensively describe the effect of the relative register - or, as it will be referred to, phase - between the screens. Figures 1a and 1b demonstrate that the latter should not be neglected. Both show a moiré free combination of the same set of three screens, but with different relative phases with regard to each other. The "micro moiré", commonly referred to as the rosette is totally different. The one in Figure 1a is called "clear centered", as opposed to the other one which is called "dot centered".

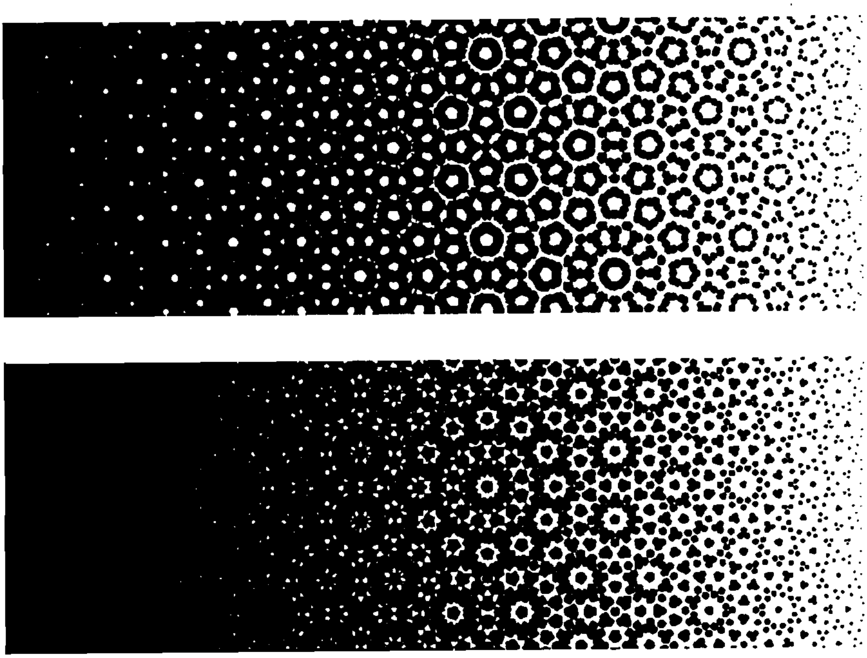

Figure 1: Two moiré free combinations of the same set of three screens. The upper set (a) shows a clear centered rosette, while the lower one (b) has a dot centered rosette.

In this paper the morphology of "the rosette" in four color printing is investigated as well as its effect on color balance. Such a study requires to look at both the amplitude <u>and</u> phase spectra of the Fourier transform of the traditional halftoning system. Since an elaborate mathematical elaboration would require the use of long and heavy expressions which would make this paper difficult to read, the text will concentrate on the qualitative interpretation of the findings. The complete derivation, however, will be published at a later date.

A first and second part of this text concentrate on the relation in traditional halftoning between the relative register of the halftone screens and the visual appearance of the "rosette". In a third part, the color balance with the clear and dot centered rosette structure is compared with the predictions as made by the Neugebauer expressions when relative randomization of the halftone dots is assumed. Finally, an improvement is suggested in the last part of the text to the traditional four color halftoning technique that optimizes color balance.

PART 1

Fourier analysis of a single screen

The Fourier spectrum of an amplitude modulated dot screen with origin (x_0, y_0), with a period T, an angle α, and square halftone dots with width w can be calculated and is expressed by the summation of terms which are the product of three functions:

$$\mathcal{F}(f_x, f_y, T, \alpha, w, x_0, y_0) = \frac{w^2}{T^2} \sum_{n=-\infty}^{n=+\infty} \sum_{m=-\infty}^{m=+\infty} F_1(f_x, f_y, T, \alpha, n, m) * F_2(T, w, n, m) * F_3(n, m, x_0, y_0)$$

in which:

$$F_1(f_x, f_y, T, \alpha, n, m) = \delta\left(f_x - \frac{n\cos\alpha}{T} + \frac{m\sin\alpha}{T}\right) * \delta\left(f_y - \frac{n\sin\alpha}{T} - \frac{m\cos\alpha}{T}\right)$$

$$F_2(T, w, n, m) = \mathrm{sinc}\left(\frac{w(n+m)}{T\sqrt{2}}, \frac{w(m-n)}{T\sqrt{2}}\right)$$

$$F_3(n, m, x_0, y_0) = e^{-2\pi i(nx_0 + my_0)}$$

The significance of these three functions is easily interpreted:

1. The first function $F_1(f_x, f_y, T, \alpha, n, m)$ indicates that the only frequencies at which the Fourier spectrum has a non zero value are the DC component, the two orthogonal fundamental screen frequencies, their harmonics and all the combinations of the latter. From this follows that the power spectrum of a grid of halftone dots consists of a grid itself.

2. The second function $F_2(T, w, n, m)$ predicts the amplitude of the different frequency components. The amplitude of the "DC component" (corresponding to n=m=0) is equal to w^2 / T^2. The amplitude of the other frequency components diminishes as they

contain higher order harmonics. The exact nature of the second function depends on the actual dot shape, and only corresponds to the "sinc" function if square dots are used. As this becomes important in the remainder of the text, it should be noted here that the width of the "lobs" of the sinc function becomes wider as the tone scale approaches its extremes. More specifically does this mean that the amplitude of the harmonics decreases less quickly (does even not decrease at all in the limit case) and reverses its sign less often when the tone value of the screen approaches 0% or 100%. As the rosette pattern will turn out to be a product of interactions between these harmonics, it will be shown later on that this effect is responsible for an increased or decreased visibility of the rosette patterns near the ends of the tone scale.

3. The third function $F_3(n,m,x_0,y_0)$ is always equal to a complex number of which the modulus is equal to 1. This term contains only information on the phase of the corresponding frequency component. As the formula shows, the phase of an individual frequency component changes linearly with the degree of the harmonics it contains. This is to be expected since it also means that the *relative phase relation* between the different frequency components, and hence the visual appearance of the screen, is not affected by a shift of the origin of the screen.

A representation of the Fourier transform is depicted in Figure 2. The dots in the drawing represent the position in the two dimensional Fourier domain where energy is present. With every dot a complex number is associated that corresponds with one of the terms in the Fourier expression above. The size of the dots is to be interpreted as a measure for the modulus of the corresponding term. There are two special cases in which the phase factor $F_3()$, normally a complex number, becomes real: When $x_0 = y_0 = 0$, $F_3()$ becomes $\underline{+1}$ for <u>all</u> the frequency components. However when $x_0 = y_0 = 0.5$, $F_3()$ has a value that <u>alternates</u> between $\underline{+1}$ and $\underline{-1}$ depending on the order of the harmonic component. An attempt to represent the distinction between these two special cases was made by making all the dots in Figure 2a black, while they alternate between white and black on Figure 2b.

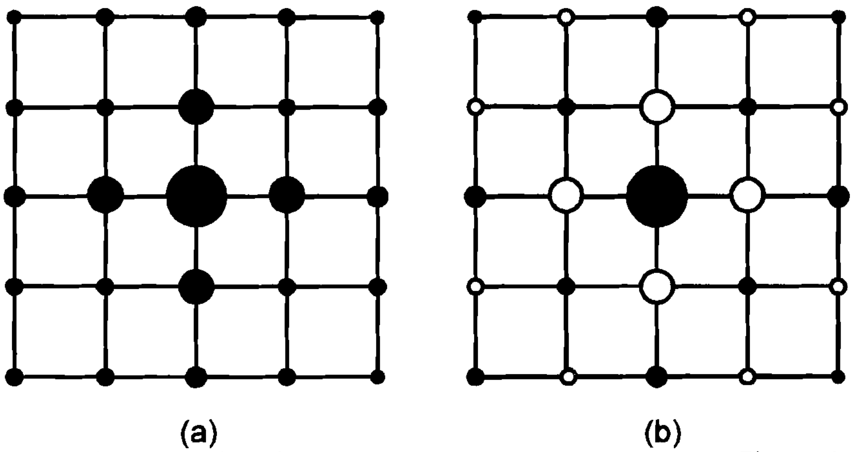

Figure 2: Representation of the Fourier spectrum of a single screen. The origin of the screen in (b) was shifted by half a period along its x and y directions.

The fact that shifting the origin of a halftone screen causes the sign of certain harmonics to change may look strange at first. The same however happens in the one dimensional Fourier domain where a shift of the origin over half a period of a signal with period T results in a change of the sign of the even harmonics. This is illustrated by means of Figure 3.

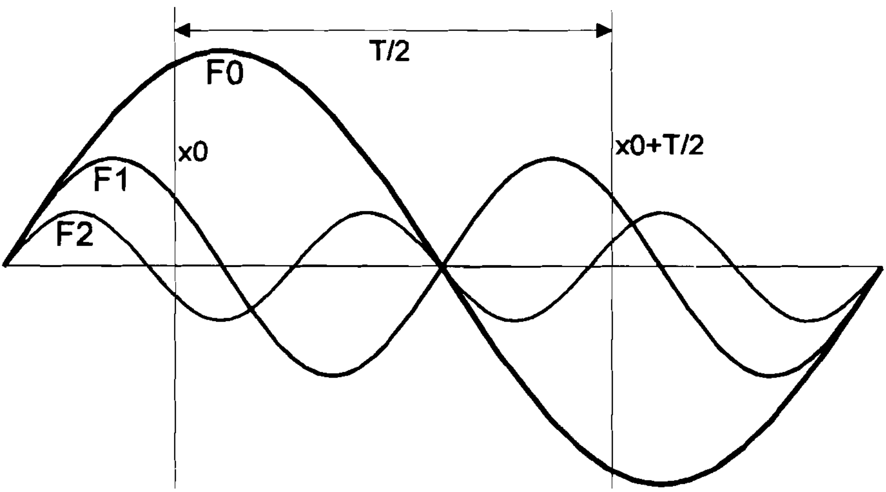

Figure 3: A shift of the origin x_0 over half a period T/2 causes the sign of the even harmonics of a periodic signal to reverse.

Fourier analysis of three cumulative screens

The four color printing process is based on the use of four screens that absorb different parts of the *visual* spectrum. Because of the "side absorptions", however, there is a substantial part of the *visual* spectrum in which the absorptions of the black, cyan and magenta inks overlap. The combined absorption at every wavelength in this part of the *visual* spectrum equals the product of the absorptions of the individual screens. The theory of Fourier analysis teaches that in such a case the *Fourier* spectrum of the combined absorption is obtained by convoluting the *Fourier* spectra of the individual absorptions. This is how the following expression was obtained of the *Fourier* spectrum of the combination of three screens with a period T, angles $\alpha_1, \alpha_2, \alpha_3$, dotwidths w_1, w_2, w_3 and positions $(x_1, y_1), (x_2, y_2), (x_3, y_3)$:

$$\mathcal{F}(f_x, f_y, T, \alpha_1, \alpha_2, \alpha_3, w_1, w_2, w_3, x_1, y_1, x_2, y_2, x_3, y_3) =$$

$$\mathcal{F}(f_x, f_y, T, \alpha_1, w_1, x_1, y_1) \otimes \mathcal{F}(f_x, f_y, T, \alpha_2, w_2, x_2, y_2) \otimes \mathcal{F}(f_x, f_y, T, \alpha_3, w_3, x_3, y_3)$$

in which the symbol $\otimes$ refers to the "convolution operation".

Working out this expression leads to:

$$\mathcal{F}() = \frac{w_1^2}{T^2} * \frac{w_2^2}{T^2} * \frac{w_3^2}{T^2} \sum_{n,m,k,l,p,q=-\infty}^{n,m,k,l,p,q=+\infty} G_1() * G_2() * G_3()$$

in which:

$$G_1() = F_1(f_x, f_y, T, \alpha_1, n, m) \otimes F_1(f_x, f_y, T, \alpha_2, k, l) \otimes F_1(f_x, f_y, T, \alpha_3, p, q)$$

$$G_2() = F_2(T, w_1, n, m) * F_2(T, w_2, k, l) * F_2(T, w_3, p, q)$$

$$G_3() = e^{-2\pi i(nx_1 + my_1 + kx_2 + ly_2 + px_3 + qy_3)}$$

The three functions $G_1(), G_2(), G_3()$ can be interpreted in a similar way as for a single screen:

1. The first function $G_1()$ indicates that the combined Fourier spectrum is non zero at all possible vectorial additions of the frequency vectors in the original spectra.

2. The second function $G_2()$ expresses the amplitude of each of these frequency components. It is important to remark that the amplitude of the DC component in this case is not only determined by the DC component of the individual screens (n=m=k=l=p=q=0), but also

by the amplitude (and phase...) of some of the other components! In fact, all the components from the three screens that, when vectorially added together, form "closed triangles" will contribute to the DC component of the combined halftone.

3. The third function $G_3()$ once again consists of a complex number that represents the phase of the combined frequency component. As the expression shows, this phase depends on the position of the origin of all three screens of which the combined halftone consists. The phase is <u>not</u> a linear function of the position coordinates of the individual screens, and this already suggests that the way that the frequency components add up, and hence the visual appearance of the combined halftone, will be affected by register.

As is shown in the next part, it is the latter property that alters the appearance of the rosette as a function of relative shifts between the screens.

PART 2

Visual appearance of dot versus clear centered rosette

Before proceeding to the actual discussion, a number of conventions and a consistent nomenclature have to be agreed on. In what follows, the values 0 and 1 refer to the darkest and lightest tone values respectively. Areas where the halftone screen is "white" are called "white halfdots", as opposed to the "black halfdots". The power spectrum in the Fourier domain is to be interpreted as the periodic presence of white halfdots. A relative coordinate system is placed with its origin at the center of a white halfdot.

One of the strongest interference patterns appears when the values $(x_1, y_1), (x_2, y_2), (x_3, y_3)$ are all equal to zero. This corresponds to the situation in which the origins of all three screens are aligned with regard to each other. Except if (at least) one of the screens is completely dark, this combination will always leave the origin white, and corresponds to the *clear centered* rosette. The value of the function $G_3()$ for this case is always a positive, real number equal to 1.

A *dot centered* rosette is formed when the centers of black halfdots coincide. This configuration is obtained from the clear centered rosette by shifting the origins of each of the three screens by half a period in their respective x and y directions. The total amount of phase shift in this case of a combined frequency component is expressed by $e^{-2\pi i(0.5m+0.5n+0.5k+0.5l+0.5p+0.5q)}$. This equation can be simplified to $(-1)^{m+n+k+l+p+q}$. This value is also real, but it can be positive or negative depending on the order of the harmonics that contribute to the frequency component.

One example where a shift of the origins with half a period is implicitly performed is when the polarity of a set of screens is reversed. A clear centered rosette is transformed into a dot centered one by this operation and vice versa.

A graphical representation of a set of three screens in the clear centered configuration is shown in Figure 5. In order not to overload the drawing, only the frequency components of the individual screens are shown. The additional components that are introduced as a result of the convolution and that are responsible for the origination of the rosette structure can be thought of as all the possible vectorial sum and differences between the original components. There are places (indicated by the circles) where these original components are quite close together, and these combinations will give rise to interactions with frequencies that are much lower - and therefore much more visible - than the frequencies of the halftone screens themselves.

The clear and dot centered rosette configurations have some quite opposite aspects that are now discussed.

The clear centered rosette is more visible in the shadows.

It was mentioned already that in the Fourier expansion of the combined spectrum the phasefactor $G_3()$ is equal to +1 for <u>all</u> the terms in the case of the <u>clear</u> centered rosette, while it <u>alternates</u> between +1 and -1 for the <u>dot</u> centered configuration. Especially in the shadows, where also the sign of $G_2()$ is positive for most of the terms, the summations therefore lead to higher values for the amplitude (and power) of most of the Fourier sums with the clear centered rosette, and hence to more visible structures.

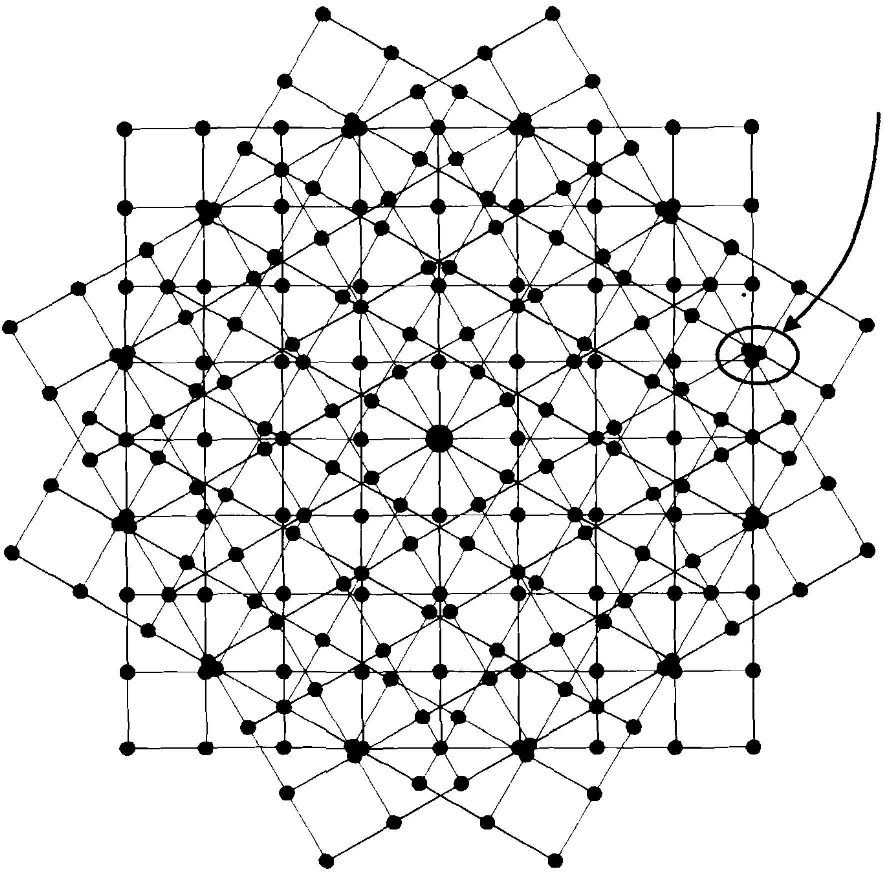

Figure 4: Fourier components of three screens in clear centered phase configuration.

This can also be understood by looking at the drawings 4 and 5. In the first drawing (corresponding to the clear centered rosette), the amplitudes of all the components in the original screens are positive, and they will add up with each other, resulting in a powerful interaction. Inversely will the positive and negative signs of the amplitudes in the second drawing sometimes cancel out each other, resulting in less powerful interactions.

One special case of this effect is that the "DC" value of the Fourier spectrum of the clear centered rosette has a higher value than for the other rosette structures. Similarly is this the case for many of the other

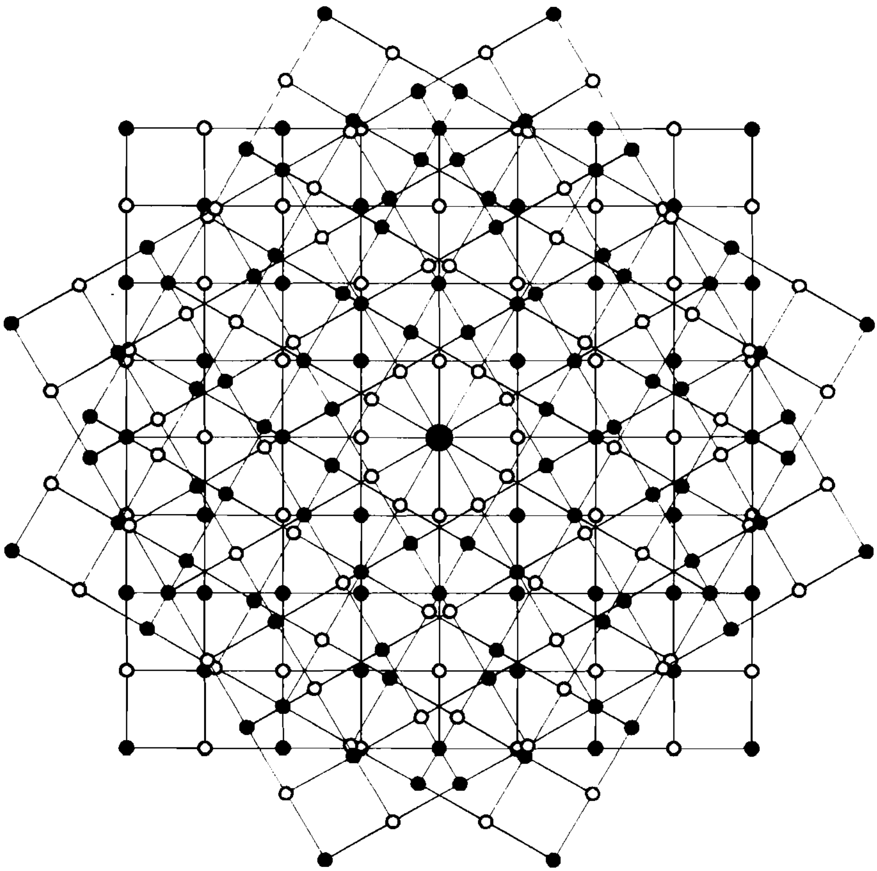

Figure 5: Fourier components of three screens in dot centered phase configuration.

components, including low frequency ones, and this explains why, in the shadows, the clear centered rosette produces the more visible structure. This effect is only supported by the fact that a correlated structure of white spaces is particularly visible in dark areas.

The dot centered rosette is more visible in the highlights.

Since the clear and dot centered rosettes are each other phase opposite, the exact opposite reasoning can be made for the black dots in the highlights: the dot centered rosette gives rise to patterns of black dots that create more objectionable structures in the highlights

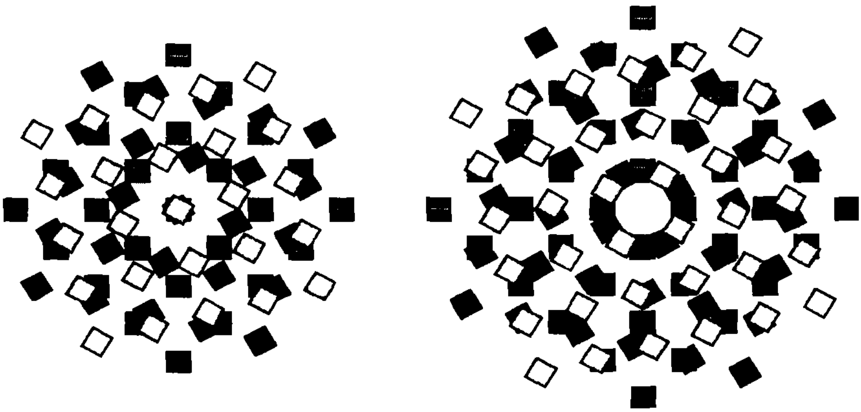

Figure 6: The smallest black structure in the highlights is larger for the dot than for the clear centered rosette.

than the clear centered rosette. An observation that supports this statement is demonstrated in Figure 6: the smallest black structure in the highlights is larger for the dot (2 times the screen period), than for the clear centered rosette($\sqrt{2}$ times the screen period) making the first one more objectionable.

There is a difference in the amount of paper covered along the "equidensity" axis.

The equidensity axis contains the colors with equal tone values for the individual screens. The difference of paper coverage using different rosette structures along this axis is illustrated by means of Figure 7. By means of a computer experiment, the percentage of non covered paper was counted along this wedge and is plotted. While the clear centered rosette leaves a significant portion of the paper uncovered up to 100%, this is not the case with the dot centered rosette.

This result can be understood in the context of the phase relations between the dots. The white portions of the paper are only saved by the three screens if they all produce white. Therefore this happens more often when these portions are in phase with regard to each other which is particularly the case for the clear centered rosette in the shadows.

It is also possible to verify that the DC component of the clear centered rosette is larger than that of the dot centered rosette by looking at the terms in the general expression that contribute to the

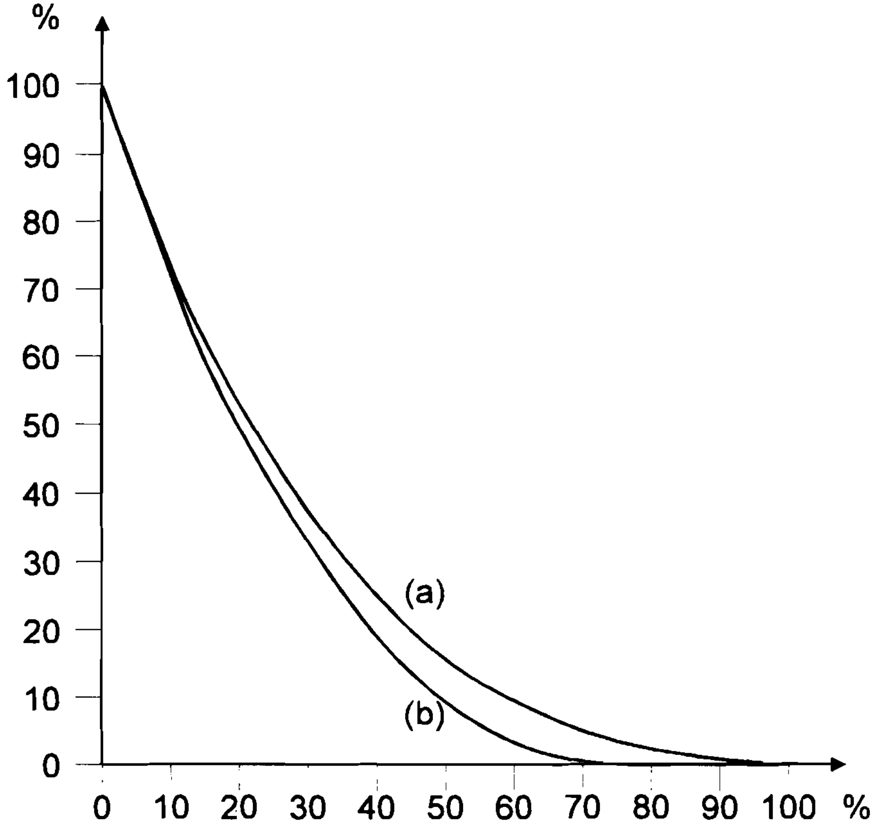

Figure 7: DC component of clear (a) and dot (b) centered rosette along equidensity axis.

"DC" term. This "DC" term consists only of these combinations of fundamentals and harmonics of which the vector sum adds up to a vector with zero length by forming a "closed triangle". It can be shown from the general expression above that the "DC" component in the case of the clear centered rosette is expressed as a function of the dot area $(w / T)^2$ of the contributing screens by the following formula:

$$\mathscr{F}(\text{"DC, equidensity, clear"}) = \frac{w^3}{T^3}\sum_{n=-\infty}^{n=\infty}\sum_{m=-\infty}^{m=\infty}\left\{\mathrm{sinc}^3\left(\frac{w(n+m)}{T\sqrt{2}},\frac{w(m-n)}{T\sqrt{2}}\right)\right\}$$

and for the dot centered rosette by:

$$\mathscr{F}(\text{"DC, equidensity, dot"}) = \frac{w^3}{T^3}\sum_{n=-\infty}^{n=\infty}\sum_{m=-\infty}^{m=\infty}\left\{(-1)^{(n+m)}\,\mathrm{sinc}^3\left(\frac{w(n+m)}{T\sqrt{2}},\frac{w(m-n)}{T\sqrt{2}}\right)\right\}$$

The evaluation of the above expressions was found to be in exact agreement with the results from the computer experiment shown in Figure 7.

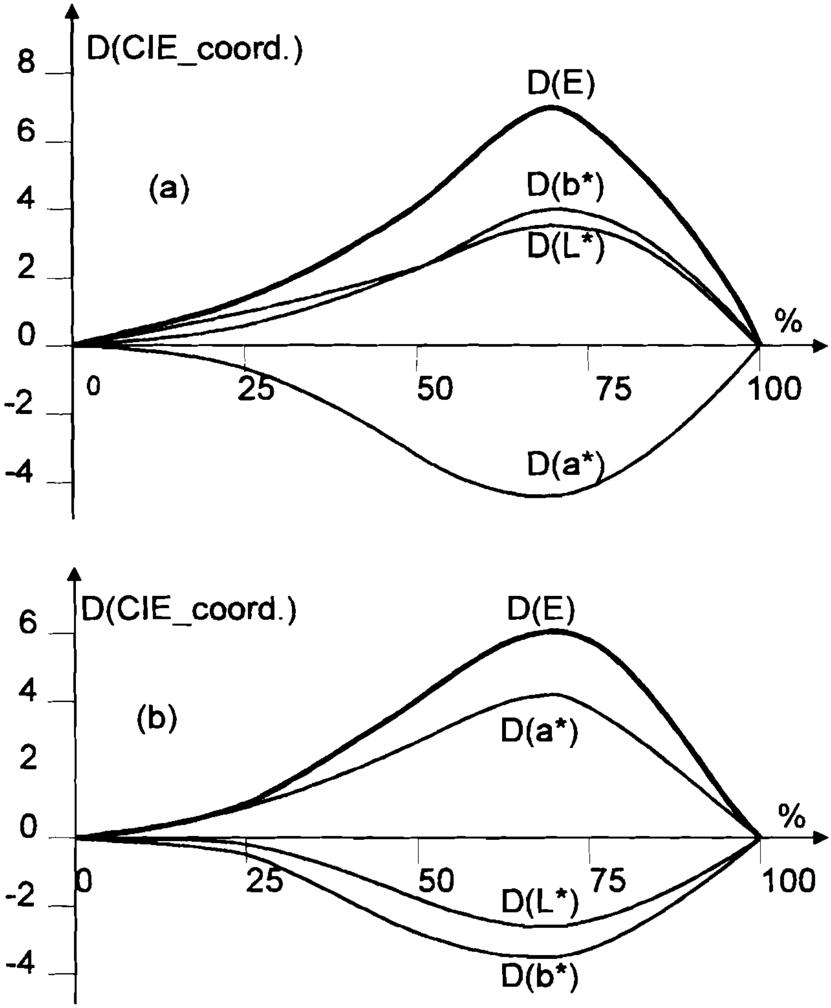

Figure 8: The Lab values along the equidensity axis differ from the colors as predicted for stochastic screening with both the clear (a) and dot (b) centered rosette.

Different rosette structures require different color calibrations

It has been observed in practical situations that the rendering of neutral colors tends to be too magenta if a dot centered rosette is used and too green with a clear centered rosette. In the presence of moiré, a low frequent component oscillates between these two colors. This means that, in theory, the calibration of a color separation process is specific for a given type of rosette, and is not by default compatible with the calibration for a stochastic screening method. A quantitative discussion of the difference in color is discussed in what follows.

PART 3

Color shift with dot versus clear centered rosette

The printing with three inks and three halftone screens results theoretically in 8 possible combinations of ink overlap. The Neugebauer expressions predict the resulting color as a linear function of the colors of these combinations. The Neugebauer equation for the X tristimulus value in a three color printing process is:

$$X(c_1,c_2,c_3) = a_w X_w + a_1 X_1 + a_2 X_2 + a_3 X_3 + a_{23} X_{23} + a_{13} X_{13} + a_{12} X_{12} + a_{123} X_{123}$$

The terms X_{ijk} are the X tristimulus values of the corresponding overprints. The Neugebauer expression for the Y and Z tristimulus values are obtained by replacing the the X tristimulus values by the corresponding Y and Z values respectively. If it is assumed that the relative positions of the halftone dots is random, the Neugebauer coefficients a_{nm} can be calculated from the Demichel equations that predict the fraction of each combination of the three inks as a function of their respective dot percentages c_1, c_2 and c_3, and this leads to the Neugeauer equations in their most often encountered form:

$$a_w = (1 - c_1) * (1 - c_2) * (1 - c_3)$$
$$a_1 = (c_1) * (1 - c_2) * (1 - c_3)$$
$$a_2 = (1 - c_1) * (c_2) * (1 - c_3)$$
$$a_3 = (1 - c_1) * (1 - c_2) * (c_3)$$
$$a_{23} = (1 - c_1) * (c_2) * (c_3)$$
$$a_{13} = (c_1) * (1 - c_2) * (c_3)$$
$$a_{12} = (c_1) * (c_2) * (1 - c_3)$$
$$a_{123} = (c_1) * (c_2) * (c_3)$$

Instead of assuming, like in the Demichel equations, that the dot positions of the three colors are randomized, it is also possible to calculate these coefficients by counting in a computer experiment the fractions of the different ink combinations obtained with the conventional screening system using either the clear or dot centered rosette configuration.These coefficients can than be used to calculate the CIE XYZ or Lab colors along the equidensity axis using the black, cyan and magenta primaries of, for example, Agfa's proofing system. Figure 8 shows how the Lab values of these colors as predicted for the clear or dot centered rosette structures differ from the colors calculated from the Demichel equations. Both plots also show the total visual difference Delta_E between the predictions. The largest color differences between the colors calculated from the reference colors and the colors obtained with clear or dot centered rosette are found around the 70% dot value on the equidensity scale. Around this value, the clear centered rosette produces a color that is too light (L^* too high) and too green (a^* too low). The dot centered rosette produces colors that are too dark (L^* too low), and too magenta (a^* too high).

The values that are shown in Figure 7 and 8 should certainly be seen as maximum deviations that only occur for the purely dot or clear centered rosette, and for equidensity colors. As soon as one of these conditions is not met, the deviations become smaller. In addition does the "softness" of printed halftone dots explain why the color deviations in practical situations are usually smaller than what is predicted by the curves in these figures.

PART 4

Randomizing dot centers reduces rosette structure.

It should be clear from the previous explanation that the visibility of rosette patterns is particularly an issue at the extremes of the tone scales, where the amplitude and phase of the harmonics that are responsible for their origination change least quickly with their order and therefore amplify or cancel out each other most strongly. A solution that was found to be effective in reducing the visibility of these patterns is the addition of a random phase vector to the position of each dot center. Such a random vector breaks up the phase coherence that exists between the harmonics that are normally responsible for the rosette formation. A refinement of this technique is applied in the Agfa Balanced Screening Technology and consists of

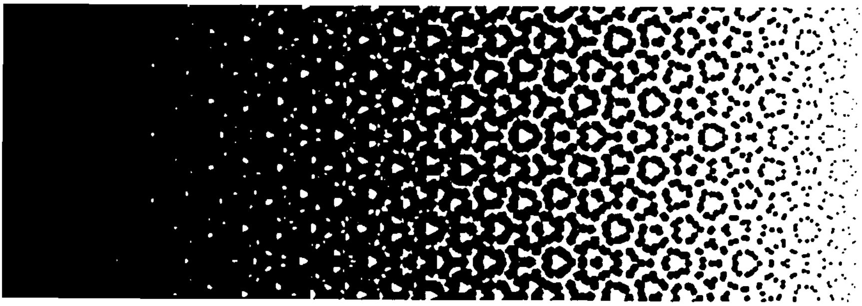

Figure 9: Degrade with variable rosette structure.

changing the modulus of this vector in a tone dependent fashion, thereby achieving the suppression of the rosette structures in combination with an optimal signal to noise ratio of the halftones.

Tone dependent phase modulation

In the above discussion, the effect of the rosette has been discussed on both the visibility of its structure, and on color balance. It was shown that, with regard to the first criterion, the clear centered performs the best in the highlights and the dot centered rosette in the shadows, but

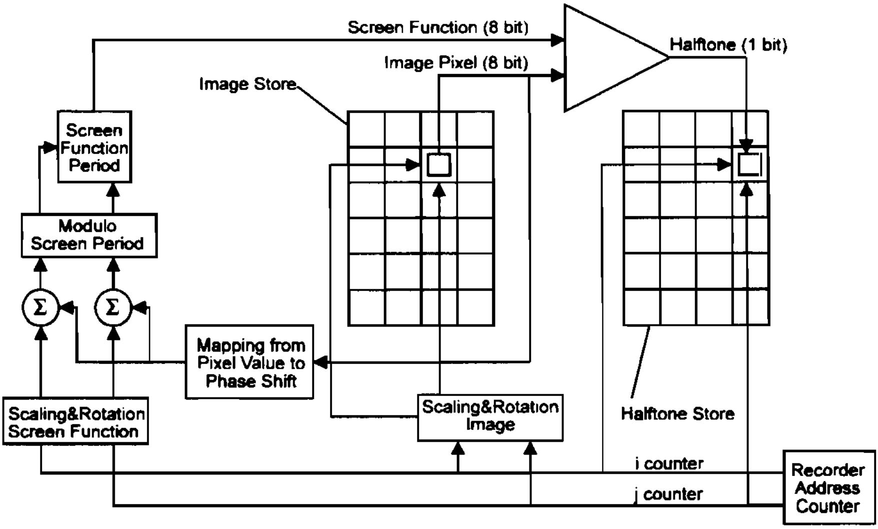

Figure 10: Modified halftone screen generator includes tone dependent phase modulation.

that both configurations produce opposite color deviations on the neutral scale in the midtones. These observations have lead to research the possibilities of using a variable rosette structure. By varying the rosette structure in a tone dependent fashion, its visibility could be optimized in both the highlights and the shadows, and in addition could the neutral balance be improved. Figure 9 gives an example of a degrade that was generated this way.

Such a tone dependent rosette structure is for example achieved by modulating the phase of each of the individual screens as a function of the tone value. In order to obtain the two extreme opposite rosette configurations for the minimum and maximum dot percentages, a maximum shift of the center of (0.5,0.5) along the internal diagonal of each of the three screens is required. It can be shown that, by optimizing the amount of phase shift as a function of tone in between these two extremes, it is possible in theory to obtain exactly the same color balance as predicted by the Demichel equations. This means that now the same color balance can be obtained in combination with either a moiré free conventional or a stochastic screening method. Experimental data, based on both computer experiments and measurements on real samples have fully supported this important conclusion.

Several methods exist to modulate the phase of a screen as a function of tone. In one method, a tone dependent bias is added to the coordinate values that address the screen function in the screen

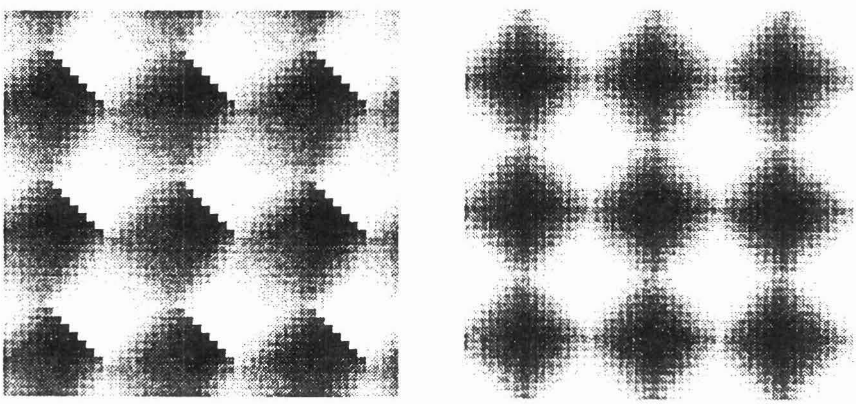

Figure 11: Spotfunction with (a) and without (b) built in phase shift.

generator. Figure 10 depicts a halftone generator that was modified according to this principle. In a second method, the tone dependent phase shift is built in the screen function itself. Figure 11 shows a representation of a screen function with and without such a tone dependent phase shift. The latter method has the advantage that it can be implemented on most existing screening hardware and software devices, including all PostScript RIP's.

CONCLUSION

In order to study the visibility of rosette patterns, it is necessary to include the effects of the harmonics of halftone screens. The harmonics all add up in the shadows with the clear centered rosette and in the highlights with the dot centered rosette, producing objectionable patterning. Both rosette types produce opposite color deviations in the mid and three quarter tones compared with stochastic screening. By making the rosette structure tone dependent, its visibility can be minimized in both the highlights and the shadows, and the color balance can be made identical to the color balance obtained with stochastic screening. A tone dependent rosette is achieved by modulating the phase of the halftone screens as a function of tone.

References

(1) Screening Techniques, Moiré in Four Color Printing, Paul A. Delabastita, TAGA proceedings 1992. (44-65)

(2) Linear systems, Fourier Transforms and Optics, Jack D. Gaskill, @1978 by John Wiley & Sons, Inc.