

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
13 April 2006 (13.04.2006)

PCT

(10) International Publication Number
WO 2006/039368 A2

(51) International Patent Classification:
G01R 7/00 (2006.01)

(21) International Application Number:
PCT/US2005/034926

(22) International Filing Date:
29 September 2005 (29.09.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/615,038 1 October 2004 (01.10.2004) US

(71) Applicant (for all designated States except US): **INDIANA UNIVERSITY RESEARCH AND TECHNOLOGY CORPORATION** [US/US]; 351 West 10th Street, Indianapolis, IN 46202 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **RUBINSTEIN, Jacob** [IL/US]; 2631 E. Poplar Drive, Bloomington, IN 47401 (US). **WOLANSKY, Gershon** [IL/IL]; 21 Neve Shaanan Street, Jerusalem (IL).

(74) Agent: **FORMAN, Alexander, D.**; Ice Miller, One American Square, Box 82001, Indianapolis, IN 46282-0200 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

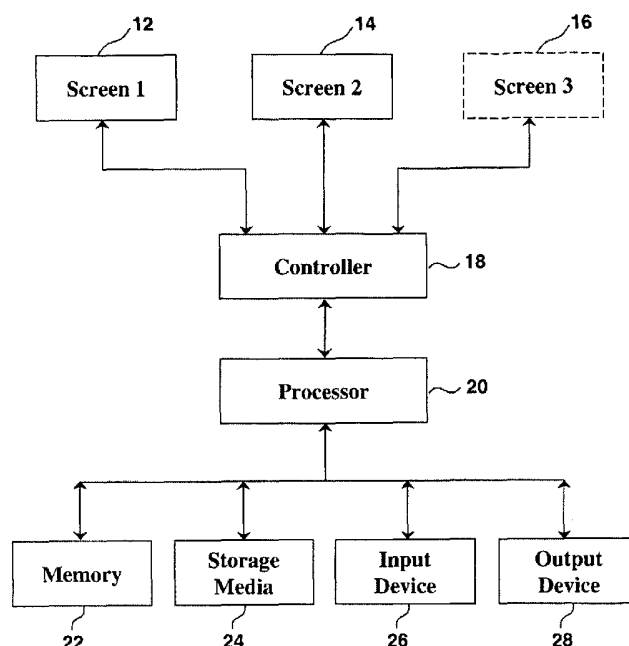
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: PHASE DETERMINATION SYSTEM AND METHOD



(57) Abstract: The present invention relates to the field of optics, and, in particular to a system and method for determining the phase of a wave. The system and method of the present invention calculates the phase of any type of wave based on the intensity of the wave. In particular, the system and method provides a means for finding the ray mapping between two surfaces in space from information on the wave's intensity measured at those two surfaces.

PHASE DETERMINATION SYSTEM AND METHOD

CROSS REFERENCE TO RELATED APPLICATION

This application claims the benefit of United States Provisional Patent Application No. 60/615,038 filed October 1, 2004.

5 Field of the Invention

The present invention relates to the field of optics, and, in particular to a system and method for determining the phase of a wave.

Background of the Invention

10 A central problem in the field of optics is determining the phase of a wave. The need to find the phase arises in a variety of applications including adaptive optics, astronomy and ophthalmic optics. The problem of phase determination is particularly difficult when the phase is not very close to being planar or spherical, and therefore interferometry methods are difficult to apply. A variety of devices, called phase sensors, have been developed to measure the phase of a wave.

15 Such phase sensors are useful in a variety of applications. For example, in ophthalmology, aberrometers are devices that utilize a phase sensor to measure the phase of a wave to characterize a subject's vision. An aberrometer sends a narrow beam of light from outside the eye into the eye to interact with the retina. When the beam interacts with the retina, a light source is created at a point in the center of the retina. This light source expands as the
20 waves travel outside the eye from the retina. During this travel, the wavefront, (as is consistent with common practice in the art, the terms "phase" and "wavefront" are used interchangeably in this application for convenience, even though technically the term "phase" refers to the fundamental object that characterizes the wave and the term "wavefront" refers to the surface in

space where the phase is constant) that is initially spherical near the retinal source point, is refracted and distorted. If the eye is perfect, the wavefront emerging from the eye would be exactly planar, implying that a distant object is exactly imaged at the retina. Or, if the subject is a "perfect" myope (near-sighted), the emerging wavefront would be spherical about some point
5 outside the eye, and the subject would have perfect focusing ability with respect to this point. In practice, no subject has a perfect eye, however, and, thus, the emerging wavefront has a nonplanar and nonspherical form.

Aberrometers contain a phase sensor to measure the refracted wavefront of light outside the eye. The measured phase is used to characterize the subject's vision. For example, from the
10 phase measurement of an aberrometer, one can, in principle, determine the prescription for the subject to improve the subject's vision. In practice, the measured wavefront is often supplied to the operator of the aberrometer, not as a function, but as an array of numbers referred to as "aberrations". The first few aberrations are related to the subject's eye prescription. Additional numbers, referred to as "higher aberrations", are supplied by the aberrometer, but these higher
15 aberrations are generally not measured in most eye examinations.

Aberrometers, or like devices, have many potential applications. In refractive surgery (i.e., Lasik surgery), where a surgeon ablates the cornea to correct a subject's vision, the ablation process might create unwanted aberrations that were not present before the surgery. Therefore, surgeons often use an aberrometer for quality control purposes. Another promising potential
20 application of aberrometers is its use in optometrist's office as a faster and reliable means for determining a subject's need for visual correction, and even for better design of visual aid for the subject.

A widely used general phase sensor in aberrometers is the Hartmann-Shack device. It consists of an array of lenslets that convert an incoming beam into spots of light on a detection screen. This sensor has a number of drawbacks: the resolution is limited by the size of lenslets; the location of the spot centroids is hard to determine accurately; and the transformation from the location of the centroids of the spots to the phase gradient is only approximate. It is desired to provide a system and a method for determining the phase of a wave that does not suffer from these drawbacks.

Another means that can be used by a system with phase sensors to determine the phase of a wave is the Fermat principle, which is one of the pillars of optics. The principle lies at the foundations of geometrical optics, where it provides a theoretical and computational tool to find ray trajectories, and hence the phase of a wave. Fermat postulated that a light ray travels between two specified points so as to minimize the action $\int n dl$, where n is the refraction index of the medium. The Fermat principle requires that one be given the terminal points of a ray. Because the terminal points of a ray are difficult to obtain and not always available, it is also desired to provide a system and method for determining phase without requiring the terminal points of a ray.

Unlike the terminal points of a ray, it is relatively easy to measure a wave's intensity. It is therefore tempting to seek methods for finding the phase from intensity measurements. Indeed, such a phase sensor was proposed in the "Deterministic Phase Retrieval: A Green's Function Solution" article by M.R. Teague that appeared in the 1983 Journal of the Optical Society of America, Volume 73 at pages 1434-1441 (the "Teague Article"), which is incorporated herein by reference. To explain the idea behind such sensors (sometimes called

"curvature sensors"), consider a wave u in the Fresnel regime where the wave equation in a homogeneous medium is given by:

$$-i \frac{\partial u}{\partial z} = ku + \frac{1}{2k} \Delta u. \quad (1)$$

Here, z is the main direction of propagation, u is the wave function, k is the wave number, and

5 ∇ and Δ denote, respectively, the gradient and laplacian operators in the plane orthogonal to z .

Writing $u = Ae^{ik\phi}$, we obtain for the real and imaginary parts of equation (1):

$$-\frac{\partial A}{\partial z} = \frac{1}{2A} \nabla(A^2 \nabla \phi), \quad (2)$$

$$\frac{\partial \phi}{\partial z} = 1 + \frac{1}{k^2 A} \Delta A - \frac{1}{2} |\nabla \phi|^2. \quad (3)$$

10 The first equation can be written more conveniently as an equation for the intensity $I = A^2$:

$$-\frac{\partial I}{\partial z} = \nabla \cdot (I \nabla \phi). \quad (4)$$

Equation (4) is called the *Transport of Intensity Equation* ("TIE"). The Teague Article points out that equation (4) can be thought of as an elliptic partial differential equation for the phase ϕ in terms of the intensity I . Thus, the Teague Article considered equation (4) over some domain D

15 in a plane $z = z_0$, and solved it under prescribed boundary conditions (the Dirichlet problem).

We note that strictly speaking the phase is $k\phi$, but, herein we shall refer to ϕ alone also as the phase.

The difficulty with the Teague Article's method is that the values of the phase at the boundary ∂D are not easy to measure. A number of algorithms related to equation (4) that
20 attempted to resolve this issue were suggested, but none of them are fully satisfactory. Another

drawback of the approach based on solving equation (4) is that it requires the computation of the derivative of the intensity. It is well-known that differentiation of data may amplify measurement errors.

Notice that the TIE is only one half of the Fresnel equation. Clearly, a proper solution
5 must satisfy the other half of system (3) and (4) as well. This raises the following question:
whether the intensity I measured at *two* planes $z = Z_1, z = Z_2$ can be used to determine the
phase by considering jointly equations (4) and (3). The measurement of the intensity at two
planes is also required for the TIE itself, because one needs to find, not only the intensity I , but
also its derivative I_z . Computing this derivative requires one to measure the intensity at two
10 near-by planes. In the present invention, however, the two planes can be arbitrarily located and
do not need to be two near-by planes.

A partial answer to this question of whether intensity measured at two arbitrary planes
could be used to determine phase was given by the "Wavefront Sensing From Defocused Images
By use of Wavefront Slopes" article by M.A. van Dam and R.G. Lane that appeared in the 2002
15 Applied Optics, Volume 41 at pages 5497-502 (the "Lane Article"), which is incorporated herein
by reference. The Lane Article realized that if the wave, confined to both observation planes,
depends only on one variable, and if the rays do not intersect, one can order the initial and
terminal points of the rays on the two respective screens such that all successive pairs of rays
hold between them the same amount of total intensity. Once the ray endpoints are known, one
20 can determine the phase slopes, and from the slopes the phase itself. The Lane Article asserts
that it is not possible to extend the method to find such a ray mapping between points in a
general two-dimensional setting. It is therefore desired to provide a system and method for

determination of a phase of a general wave based on intensities measured at arbitrarily located planes and that overcomes all the shortcomings of the systems and methods described above.

Brief Summary of the Invention

The present invention relates to the field of optics, and in particular, to a system and
5 method for determining the phase of a wave. For example, one device that can be made in
accordance with the present invention comprises a first screen and a second screen located in two
separate positions, that may or may not be in arbitrary positions, in order to measure a first
intensity and a second intensity of a wave at the two separate positions. The first and second
screens are each operatively connected to a processor, such as a computer, so that the measured
10 first and second intensities can be transferred to the processor. The processor is able to construct
the optimal mapping of the rays of the wave from the first and second measured intensities and
to calculate the phase of the wave from the constructed optimal mapping.

An exemplary method of determining the phase of the wave utilizes this device to
measure a first intensity of the wave with the first screen and a second intensity of the wave with
15 the second screen. The method constructs the optimal mapping of the rays of the wave with the
processor from the first and second intensities and then calculates the phase of the wave with the
processor from the constructed optimal mapping. As explained in more detail below, the
processor of this exemplary device and method can construct the optimal mapping by utilizing
any number of algorithms, including but not limited to, a linear programming optimization or a
20 steepest descent flow.

Any number of screens can be operatively connected to the processor and can be used to
measure any number of intensities. The processor can construct the optimal mapping based on
all the measured intensities provided by the screens. Moreover, the phase of any type of wave

can be determined utilizing these devices and methods and no assumption is made about the type of wave that is being measured, which is significant because other phase measuring devices are based on the assumption that the wave is paraxial. For example, the methods and devices of the present invention can be used to determine the phase of a paraxial, non-paraxial, symmetric, non-symmetric, spherical or non-spherical wave. The processor can comprise any type of device capable of making the necessary calculations, including but not limited to, a computer or microprocessor.

Brief Description of the Drawings

Fig. 1 shows a schematic diagram of a wavefront propagating toward the first and second screens of one embodiment of the present invention; and

Fig. 2 shows a block diagram of one embodiment of a system according to the present invention.

Detailed Description of the Invention

The present invention relates to a system and method for determining a wave's phase. In particular, one embodiment of the present invention provides a system and method for finding the ray mapping between two surfaces in space based on the intensities measured at the two surfaces, and determining phase from the ray mapping. To achieve the aforementioned system and method, a new variational principle in optics is derived and presented herein that measures intensities of the wave at two planes and uses this information to determine the phase of the wave by considering jointly equations (4) and (3). To jointly consider equations (4) and (3) for phase reconstruction, equation (3) is incorporated in the analysis of the Teague article, discussed above, by expressing the phase ϕ in the form:

$$\phi = z + \psi(x, z), \quad (5)$$

where x denotes a point in the plane R^2 , and ψ is the perturbation of the phase about the planar term z . By substituting equation (5) into equation (3), one obtains for $\psi(x, z)$:

$$\frac{\partial \psi}{\partial z} + \frac{1}{2} |\nabla \psi|^2 = \frac{1}{2k^2 A} \Delta A. \quad (6)$$

In the small wavelength approximation, neglect the term on the right hand side and replace
5 equation (6) by

$$\frac{\partial \psi}{\partial z} + \frac{1}{2} |\nabla \psi|^2 = 0. \quad (7)$$

The above discussion on the phase reconstruction is limited to homogeneous media. The
variational principle further derived below, however, is applicable to arbitrary media. When the
10 refraction index n is not constant, the term $\frac{1}{2}(n^2 - 1)$ needs to be included in the right hand side
of equation (6). Thus, the optical problem considered consists of the following equations and
boundary conditions:

$$\frac{\partial I}{\partial z} + \nabla \cdot (I \nabla \psi) = 0, \quad (8)$$

$$\frac{\partial \psi}{\partial z} + \frac{1}{2} |\nabla \psi|^2 = \frac{1}{2}(n^2 - 1) := P(x, z). \quad (9)$$

Here, $x \in R^2$, $z \in [Z_1, Z_2]$, $I(z = Z_1, x) = I_1(x)$, $I(z = Z_2, x) = I_2(x)$, where I_1 and I_2 are two
given intensity distributions. Equations (8) and (9) together with the side conditions are denoted
herein collectively as **problem (Op)**. In the next section it is shown that problem (Op) can be
solved by certain optimization problems.

The Variational Problem I: The Paraxial Limit

Consider two planes $P_1 : z = Z_1$ and $P_2 : z = Z_2$. Let I_1 and I_2 be two nonnegative functions given on P_1 and P_2 . Optically, the functions I_1 and I_2 are the measured intensities; mathematically, however, they can be considered as arbitrary density functions. Assume that the intensities are normalized to one, and that they have finite second moments:

$$\int I_1(x) dx = \int I_2(x) dx = 1, \quad \int x^2 I_i(x) dx < \infty \quad i = 1, 2. \quad (10)$$

Recall from geometrical optics that, if a point $x \in P_1$ is mapped by a ray into a point $y \in P_2$, if the refraction index near P_1 and P_2 is the same, and if the ray is approximately orthogonal to the planes, then the intensities are related by:

$$I_1(x) = I_2(T(x)) |J(T)|. \quad (11)$$

Here, $T(x)$ is the ray mapping from P_1 to P_2 , and $J(T)$ is the Jacobian of this mapping. The relation between transport equations such as (8) and the Jacobian of the ray mapping is well known in optics (See, e.g., Asymptotic Methods for Partial Differential Equations: The Reduced Wave Equation and Maxwell's Equations, Surveys in Applied Mathematics, Vol. 1, J.B. Keller et al. Editors, 1-82 (Plenum Press, 1993), which is incorporated herein by reference). We shall say that a mapping T satisfying the relation (11) *transports* I_1 to I_2 . Use the formal notation:

$$T_{\#} I_1 = I_2. \quad (12)$$

A first variational principle according to the present invention, denoted by **problem (Mp)**, is

Find a map $\bar{T}$ such that, $\bar{T}_\# I_1 = I_2$ and

$$M(I_1, I_2, \bar{T}) := \int Q(x, \bar{T}(x)) I_1(x) dx \leq \int Q(x, T(x)) I_1(x) dx \forall T_\# I_1 = I_2, \quad (13)$$

5 where the action Q is given by

$$Q(x, y) := Q(x, y, Z_1, Z_2) \equiv \min \int_{Z_1}^{Z_2} \left(\frac{1}{2} \left| \frac{dx}{dz} \right|^2 + P(x(z), z) \right) dz, \quad (14)$$

and where the minimization is among all orbits $x(z)$ such that $x(Z_1) = x$, $x(Z_2) = y$. The functional M defined in equation (13) will be called a weighted least action functional. Other weighted least action functionals will be defined later in the left hand sides of equations (15) and (27). As will be explained below, different optical setups call for the use of different weighted least action functionals.

In the homogeneous case ($P \equiv 0$), the action reduces to:

$$Q(x, y) = \frac{|x - y|^2}{2(Z_2 - Z_1)}.$$

In this case, the variational principle **(Mp)** becomes the *quadratic Monge problem (Mpq)*:

15 Find a map $\bar{T}$ such that, $\bar{T}_\# I_1 = I_2$ and

$$\int |\bar{T}(x) - x|^2 I_1(x) dx \leq \int |T(x) - x|^2 I_1(x) dx \forall T_\# I_1 = I_2. \quad (15)$$

It is shown below that the optimal mapping $\bar{T}$, among all constrained ray mappings, is the ray mapping of the optical problem. For this purpose, problem **(Op)** is related to problem **(Mp)** through several additional equivalent optimization problems. By introducing a new problem, denoted by **(Wp)**, it can be proven that its solution is the pair (I, ψ) that solves **(Op)**.

Theorem 1: Let $\sigma = \sigma(x, z) \geq 0$ and $v = v(x, z) \in R^2$ be solutions of the following optimization problem:

$$\inf_{\sigma, v} W(I_1, I_2; P) = \inf_{\sigma, v} \int_{Z_1}^{Z_2} \int \left(\frac{1}{2} \sigma |v|^2 + P\sigma \right) dx dz \quad (16)$$

5 subject to the constraints:

$$\frac{\partial \sigma}{\partial z} + \nabla \cdot (\sigma v) = 0, Z_1 \leq z \leq Z_2, \sigma(x, Z_i) = I_i(x), i = 1, 2. \quad (17)$$

Then,

$$\sigma = I, v = \nabla \psi, \quad (18)$$

where I and ψ solve **(Op)**.

10 It can further be shown that the infimum of the functional W equals M defined in equation (13). Let (I, ψ) be the solution of problem **(Wp)**. Use the phase ψ , i.e. the solution to the Hamilton-Jacobi equation (9), to define the following flow:

$$\frac{d\bar{x}}{dz} = \nabla \psi(\bar{x}(z), z), \bar{x}(Z_1) = x. \quad (19)$$

The flow (19) induces a mapping:

$$15 \quad T_{Z_1}^z(x) := \bar{x}(z). \quad (20)$$

Proposition 2: The mapping (20) transports I_1 to $I(x, z)$.

The main result of this section can now be stated:

Theorem 3: The mapping $T = T_{Z_1}^{Z=Z_2}$, where $T_{Z_1}^z$ is defined in (20), is the optimal mapping $\bar{T}$, i.e.:

$$\bar{T} = T_{Z_1}^{Z_2}. \quad (21)$$

In addition,

$$\inf_{\sigma, \psi} W(I_1, I_2; P) = \int Q(x, \bar{T}(x)) I_1(x) dx.$$

Theorem 1 states that the minimizing pair (I, ψ) for the functional W is a solution to the problem **(Op)**. Theorem 3 states that the flow induced by ψ generates the optimal mapping $\bar{T}$. Therefore, by solving the variational problem, one obtains complete information on the ray mapping and hence the phase ψ of problem **(Op)**.

10 In the special case of the quadratic Monge problem, corresponding to the optical setup of a homogeneous medium, the action Q is minimized by the straight line (ray) connecting x and $T(x)$. Therefore, the optimal map $\bar{T}$ in this case is given explicitly by:

$$T_{Z_1}^{Z_2}(x) = x + \nabla_x \psi(x, Z_1)(Z_2 - Z_1). \quad (22)$$

15 Given two intensity distributions, and assuming that they are related by the paraxial Fresnel equations, the variational problem **(Mp)** provides one with a theoretical and practical tool to find a phase map that connects these intensities.

The mathematical analysis is valid for the optical problem **(Op)** regardless of its origin. The function P then has the interpretation of the potential of the physical system, and the z coordinate represents time. Therefore, the variational principle **(Mp)** means that if one is given
20 the absolute value of the wave function everywhere in space at two different times Z_1 and Z_2 , one can find the phase of the wave function at all times $t \in (Z_1, Z_2)$.

The Variational Problem II: General Waves

In the above Section, it was shown that the data embedded in the intensity distributions given on two planes suffices to find the phase of a wave in the paraxial regime. This finding naturally raises the question of whether the result can be extended to general waves. In this
 5 Section, a positive answer to this question is presented. The idea is to use the Fermat action itself (in a suitable form) as the action in the optimization problem and to phrase the intensity equation as an equation for the radiance.

Start by formulating the optical setup. Consider a solution to the Helmholtz equation:

$$\Delta u + k^2 n^2(x, z)u = 0 \quad (23)$$

10 of the form $u = A(x, z)e^{ik\phi(x, z)}$. Expanding as usual in large k , one obtains the eikonal equation $(\frac{\partial \phi}{\partial z})^2 + |\nabla \phi|^2 = n^2$, for the phase ϕ , and the transport equation:

$$\partial_z (I \frac{\partial \phi}{\partial z}) + \nabla \cdot (I \nabla \phi) = 0, \quad (24)$$

for the intensity $I = A^2$. Here, as before, x denotes a point in the plane orthogonal to the z direction, and ∇ is the two dimensional gradient. Because one is not limited now to paraxial
 15 rays, there is some freedom in choosing the z axis. When considering general waves, it is more appropriate to introduce the radiance function. Therefore, define the radiance $\rho(x, z) = I \frac{\partial \phi}{\partial z} = I \sqrt{n^2 - |\nabla \phi|^2}$, and write the transport equation (24) with respect to it. Consider the case where the radiance is given on two parallel planes choose the z axis to be orthogonal to the planes.

To formulate the *optical problem (O)*, find the phase function $\phi(x, z)$ and the radiance function $\rho(x, z)$ such that ρ and ϕ satisfy:

$$\frac{\partial \rho}{\partial z} + \nabla \cdot \left(\rho \frac{\nabla \phi}{\sqrt{n^2 - |\nabla \phi|^2}} \right) = 0 \quad Z_1 < z < Z_2, \quad (25)$$

and

$$\left(\frac{\partial \phi}{\partial z} \right)^2 + |\nabla \phi|^2 = n^2, \quad Z_1 < z < Z_2, \quad (26)$$

subject to $\rho(z = Z_1, x) = \rho_1(x)$, $\rho(z = Z_2, x) = \rho_2(x)$, where ρ_1 and ρ_2 are two given radiance distributions. The question posed is whether problem (O) is solvable, and, in particular, whether it can be associated with an optimization (variational) principle.

The second variational problem, denoted by **problem (M)**, is:

Find a map $\bar{T}$ such that $\bar{T}_\# \rho_1 = \rho_2$, and

$$\int Q(x, \bar{T}(x)) \rho_1(x) dx \leq \int Q(x, T(x)) \rho_1(x) dx \quad \forall T_\# \rho_1 = \rho_2, \quad (27)$$

where the action Q is given by

$$Q(x, y) = \min \int_x^y n \, dl = \min \int_{Z_1}^{Z_2} n(x, z) \sqrt{1 + \frac{1}{2} \left| \frac{dx}{dz} \right|^2} dz, \quad (28)$$

and where the minimization is among all orbits $x(z)$ such that $x(Z_1) = x$, $x(Z_2) = y$.

The optimal mapping $\bar{T}$ is the ray mapping associated with (O). A key point in the analysis in the previous Section was the introduction of the corresponding Lagrangian equation (16). Thus, this analysis proceeds by defining an appropriate Lagrangian equation. The minimization of this Lagrangian equation is equivalent to the optical problem (O):

Theorem 5: Let $\rho = \rho(x, z) \geq 0$ and $v = v(x, z) \in R^2$ be solutions of the following optimization problem:

$$\inf_{\rho, v} W(I_1, I_2; P) := \inf_{\rho, v} \int_{Z_1}^{Z_2} \int n \rho \sqrt{1 + v^2} \, dx dz, \quad (29)$$

subject to the constraints:

$$\frac{\partial \rho}{\partial z} + \nabla \cdot (\rho v) = 0, Z_1 \leq z \leq Z_2, \rho(x, Z_i) = \rho_i(x), i = 1, 2. \quad (30)$$

Then,

$$v = \frac{\nabla \phi}{\sqrt{n^2 - |\nabla \phi|^2}}, \quad (31)$$

where ϕ solves equation (26).

Similarly to Proposition 2, the flow $T_{Z_1}^z(x) = \bar{x}(z)$, defined by:

$$\frac{d\bar{x}}{dz} = v = \frac{\nabla \phi}{\sqrt{n^2 - |\nabla \phi|^2}}, \quad \bar{x}(Z_1) = x \quad (32)$$

transports ρ_1 . Moreover, this flow induces the optimal mapping $\bar{T}$.

The different optimization problems associated with the problem (O) are now summarized:

Theorem 6:

(i) The optimal mapping $\bar{T}$ for problem (M) induces a ray mapping for the optical problem (O).

(ii)

$$\inf_{\rho, v} \int_{Z_1}^{Z_2} \int n \rho \sqrt{1 + v^2} \, dz dx = \inf_T \int Q(x, T(x)) \rho_1 dx = \sup_{\phi} \int \phi \rho_2 \, dx - \int \phi \rho_1 \, dx, \quad (33)$$

where the first integral is minimized under the constraint (30), the second integral is minimized among all maps T that transport ρ_1 to ρ_2 , and the last integral is maximized among all functions ϕ that satisfy $(\frac{\partial \phi}{\partial z})^2 + |\nabla \phi|^2 \leq n^2$.

The optical problem (O) was formulated in terms of the phase function $\rho(x, z)$. If the expression for the radiance function in terms of the intensity function I is substituted into equation (25), then, upon making use of equation (26) it is found that the intensity I also solves equation (25). Therefore, the radiance function ρ can be replaced in the presentation above by the intensity function I . This implies that the phase can be determined by either intensity or by radiance measurements.

10 General Discussion

The analysis provided in the previous Sections provides one with a useful tool/method for determining the phase of a wave from intensity measurements. The optical problems (Op) or (O) are difficult to solve because one is given information on the intensity or radiance at two separate planes, constrained by certain differential equations connecting them. The variational problems (Mp) or (M), on the other hand, provide a direct optimization problem to find the solution.

One may interpret the result as a way to give a physical meaning to the notion of a 'ray'. Rays are mathematical rather than physical entities. One cannot measure rays directly but the variational principle derived above provides a means for 'measuring rays'. Actually, intensities are measured, and then, through the mathematical procedure of finding the optimal mapping, the individual rays are identified (i.e., the optically correct family of constrained ray mapping is constructed).

Implementation in Phase Sensors

The method described above for determining the phase of a wave can be utilized by a phase sensor to calculate the phase of a wave. Generally, phase sensors are devices that measure the phase of a wave, and are useful in a variety of applications in physics and technology. Phase sensors also form an important part of adaptive optics systems. Particular applications of phase sensors emerged in recent years in ophthalmology. For example, as already discussed, an aberrometer is a device that utilizes a phase sensor to measure the shape of a wavefront generated by a light source on the retina, as the light wave exits from the eye. Most current aberrometers use a Hartmann-Shack phase sensor, whose shortcomings were discussed above.

While other types of phase sensors (i.e. curvature sensors) calculate the phase of a wave based on intensity measurements, as explained above, these types of sensors that use the TIE equation have a number of shortcomings. For example, it is not clear how to obtain boundary conditions for the TIE equation (4). In addition, a practical serious shortcoming is that, in the TIE equation, one needs to differentiate the measured intensities. Because measurement errors are inevitable, and because it is well-known that noise in data is amplified by a differentiation step, a phase sensor that utilizes the TIE equation (4) may provide inaccurate results, even if one somehow overcomes the serious obstacle of obtaining boundary conditions.

The system and method of the present invention overcome all the shortcomings mentioned herein. The present invention provides the phase at high resolution, is not limited to paraxial waves, uses intensity detection screens that need not be placed very close to each other and the measured data is not differentiated.

It is now explained how to utilize the above-described equations in a system and method to determine the phase of a wave. The first step in determine the phase of the wave is to measure

the intensity of the wave with at least two detection screens (i.e., a medium able to record and visually display information) of a measuring device, such as a phase sensor. Consider a wave propagating essentially along the z direction of a standard x, y, z axis. A measuring device detects the wave's intensity on a number of detection screens along its path. While at least two
5 screens can be used, one can also use three or more screens to improve accuracy. Any type of phase detection screen can be used, including but not limited to, CCD screens, to provide the intensity at a high resolution.

Referring now to Fig. 1, there is shown a schematic diagram of a wavefront 40
propagating toward the first and second detection screens 42 and 44 of one embodiment of the
10 present invention. As shown, the wavefront 40 will meet the first screen 42 before reaching the second screen 44. As previously described, first screen 42 and second screen 44 do not need to be located in close proximity to each other, and can be arbitrarily located in space and in different planes. The only limitation on the location of first screen 42 and second screen 44 is that they must be positioned to measure a first intensity of the wave with first screen 42 and a
15 second intensity of the wave with second screen 44. While not shown in Figure 1, more than two screens can be used to measure the intensity of the wavefront and the additional screens will be placed apart from one and another and the other screens, just as the first screen 42 and second screen 44 are placed apart from one another in Figure 1. Any additional screen will be positioned to measure an intensity that corresponds to the number of screens it represents in the
20 measuring device (e.g. a third screen will be positioned to measure a third intensity, a fourth screen will be measured to measure a fourth intensity, etc.).

The second step of calculating the phase of the wave in this embodiment is to utilize a device, such as a computer, processor, microprocessor and/or a computer chip, to analyze the

measured data and to determine the wave's phase. The computing element can use one of several algorithms that will be explained in some detail hereinafter. For simplicity, the algorithms presented are for the special (but very frequent) problem **Mpq** described above. To recall, problem **Mpq** (a.k.a. the quadratic Monge problem) consists of finding the optimal mapping $\bar{T}$ that solves the optimization problem (15). In general, the phase reconstruction performed by the computing element can be divided into two subparts. First, the optimal mapping $\bar{T}$ is constructed, and then the optimal mapping is used to construct the phase.

Referring now to Fig. 2, there is shown a block diagram of one embodiment of the system according to the present invention. In this embodiment, the system includes first screen 12, second screen 14, and, optionally, third screen 16. The system also includes controller 18, processor 20, memory 22, storage media 24, input device 26, and output device 28 all operatively connected to one another. As used herein, operatively connected includes any number of means of connecting electronics together known in the art including, but not limited to, a network, cables, wires, or wireless communication. In this embodiment, first, second and third screens 12, 14, and 16 comprise CCD screens operatively connected to and in bidirectional communication with controller 18 by means well known in the art. Controller 18 controls operations of first, second, and third screens 12, 14, and 16 (i.e., controls the provision of power to and instructs pictures to be taken on first, second, and third screens 16). Controller 18 also collects the data (image) collected on each of first, second, and third screens 12, 14, and 16. As already discussed, any number of screens can be used and connected to the controller 18, just as screens 12, 14, and 16 are connected to controller 18. Controller 18 can comprise a computer, processor, a microprocessor and/or other like device.

Processor 20 constructs the optimal mapping from the intensities measured on first, second, and third screens 12, 14, and 16. Processor 20 further computes the phase from the measured intensities. Processor 20 is operatively connected to memory 22 for execution of the program and non-permanent storage of data and calculations, to storage media 24 (i.e. a database) for storage of data (and/or results), to input device 26 (i.e., a keyboard) for operating processor 20, and output device 28 (i.e., a printer or computer screen) for output (such as printing or electronic communication) of results obtained from processor 20 by means well known in the art. Processor 20 can comprise a computer, processor, a microprocessor, and/or other like device.

It will be appreciated by those of skill in the art that controller 18 may be packaged with the screens, or separate controllers may be tightly coupled with each of the screens by means well known in the art (i.e., a network). Also, controller 18 and processor 20 may share a common processing unit, such as computer workstation. It will be further appreciated that processor 20, memory 22, storage media 24, input device 26, and output device 28 may comprise separate combinations or be included in a single unit, such as a computer workstation.

The following examples are included to demonstrate some of the possible embodiments of the invention. It should be appreciated by those of ordinary skill in the art that the techniques disclosed in the examples which follow represent techniques discovered by the inventors to function well in the practice of the invention, and thus, can be considered to constitute preferred modes for its practice. However, those of ordinary skill in the art should, in light of the present disclosure, appreciate that many changes can be made in the specific embodiments which are disclosed and still obtain a like or similar result without departing from the spirit and scope of the invention.

Examples for algorithms for constructing the optimal mapping $\bar{T}$

While the derivation in the Variational Problems I and II sections above were formulated for intensity functions defined over the entire plane, the theory is valid, of course, also for the special case where I_1 and I_2 are supported on finite domains, say D_1 and D_2 , respectively.

- 5 These domains can, for example, be associated with the system's aperture. It will be appreciated that there are many ways to estimate the domains D_1 and D_2 . For example, since most apertures are circular, one can select the domain D_1 to be a disc. One option to selecting the domain D_2 is to seek a disc in the plane P_2 such that the integral of the intensity I_2 over the disc D_2 will be the same as the integral of the intensity I_1 over the disc D_1 . Another way to select D_2 is to
- 10 image on the screen P_2 a very fine ring placed on the plane P_1 . Yet another way to determine the domain D_2 is to image on the plane P_2 a set of points that are located on the perimeter of the domain D_1 on the plane P_1 .

In principle, the goal is to solve an optimization problem as explained in detail in the previous Sections. In light of the special nature of the optimization problem at hand, some

15 specialized techniques found to be particularly useful are presented here.

I. Exemplary Method I: A linear programming optimization method

The quadratic Monge optimization problem (**Mpq**) can be converted into a linear programming problem that can then be solved by well-known techniques. This is done by associating problem (**Mpq**) with the Kantorovich minimization problem (**K**):

- 20 The idea is to consider the problem of minimizing the functional

$$K(m) = \int |x - y|^2 m(x, y) dx dy. \quad (34)$$

The minimization is over all densities m such that their marginal density is given by I_1 and I_2 , respectively, i.e.:

$$\int m(x, y) dx = I_2(y), \quad \int m(x, y) dy = I_1(x). \quad (35)$$

Recall that both x and y denote points in the plane. The problem of minimizing K under (35)

5 has a unique solution. Moreover, the minimizer m is supported exactly on the graph of the optimal mapping $\bar{T}$ defined above, namely $\bar{m}(x, y) = \delta(x - \bar{T}(y))$, where $\bar{m}$ is the optimal Kantorovich density.

To implement the Kantorovich method, the densities I_1, I_2 need to be discretized. That is, the intensities I_1, I_2 are approximated with *discrete* distributions. For this purpose, select
10 points $\{x_i\}, \{y_j\}$ according to *empirical* distributions, namely:

$$I_1 \approx \frac{1}{N} \sum_i^N \delta_{x_i}, \quad I_2 \approx \frac{1}{N} \sum_j^N \delta_{y_j}. \quad (36)$$

where $\{\delta_{x_i}, \delta_{y_j}\}$ are unit point masses, and $\sum_i m_i^{(1)} = \sum_j m_j^{(2)} = 1$. The corresponding Kantorovich problem (34) takes the form

$$\min_M \sum_{i=1}^N \sum_{j=1}^N M_{i,j} |x_i - y_j|^2, \quad (37)$$

15 where the minimum is taken on all non-negative, $N \times N$ matrices M which satisfy

$$\sum_i M_{i,j} = \sum_j M_{j,i} = 1/N.$$

This formulation has a unique minimizer which is a permutation matrix $M_{i,j} = (1/N) \delta_{i, \Pi(j)}$. The permutation defines a discrete mapping:

$$\Pi: \{1, \dots, N\} \rightarrow \{1, \dots, N\},$$

which is the solution of the discrete optimal mapping via:

$$T(x_i) = y_{\Pi(i)}.$$

The algorithm requires a selection of points in accordance with empirical distributions (36) that carry identical mass with respect to the intensities I_1 and I_2 . This can be done in a variety of ways. For instance, a simple suitable sampling method for the case where D_1 and D_2 are rectangles is proposed in the "Minimizing Flows for the Monge-Kantorovich" article by S. Angenent, S. Haker and A. Tannenbaum that appears in the 2003 SIAM Journal of Mathematics Annals No. 35, pages 61-97 (the "Angenent Article"), which is incorporated herein by reference. It is clear to one of ordinary skill in the art how to construct a similar sampling for discs and other domains.

The optimization problem (**Mpq**) (and therefore also the problem of phase determination from intensity measurements) was thus converted into a linear programming problems that is simple to implement on processor 20. The main drawback of the linear programming problems is that it requires work with functions of four variables, which implies a need for a system with a large computer memory.

II. Exemplary Method II: A steepest descent flow method

Gradient methods, such as the steepest descent flow method, are common in optimization. The idea is to define a flow that proceeds along the gradient of the Monge functional, while always preserving the constraint that the mapping transports I_1 into I_2 . A steepest descent flow for the problem (**Mpq**) was computed in the Angenent Article. To emphasize that this flow of transformations is not directly related to the optical flow, but rather, is a mathematical way for finding the optimal transporting map $\bar{T}$, introduce the flow as:

$$U = U(x, t) : R^2 \times [0, \infty) \rightarrow R^2$$

and the parameter t to denote 'time' for the flow. The flow starts with an initial mapping $U_0(x) = U(x, 0)$ that transports I_1 to I_2 . For example, the sampling (36) with trivial permutation can be used to generate an initial mapping.

In the algorithm of the Angenent Article one decomposes U at each time t into the
5 orthogonal sum of a gradient and a divergence-free vector field:

$$U(x, t) = \nabla p(x, t) + v(x, t), \quad \nabla \cdot v = 0. \quad (38)$$

When dealing with bounded domains, the requirement that v is divergence-free must be supplemented with the boundary behavior of v . The natural boundary condition is that the vector field v has no normal component at the boundary. Together, the requirements on v
10 imply that p must solve the following Poisson-Neumann problem:

$$\Delta p = \nabla \cdot U \quad x \in D_1, \quad \frac{\partial p}{\partial \nu} = U \cdot \nu \quad x \in \partial D_1, \quad (39)$$

where ν is the outward normal to D_1 . Equivalently, the potential p can be characterized as the function that minimizes the functional:

$$F(p) = \int |U - \nabla p|^2 dx. \quad (40)$$

15 The evolution of U is driven by:

$$\frac{\partial U}{\partial t} + \frac{1}{I_1} v \cdot \nabla U = 0. \quad (41)$$

The evolution problem (38), (39), (41) is referred to herein as the AHT flow.

An alternative steepest descent flow can be formulated by decomposing U differently. For example, one can replace the projection (38) by the orthogonal projection:

$$20 \quad U(x, t) = \nabla q(x, t) + w(x, t), \quad \nabla \cdot (I_1 w) = 0. \quad (42)$$

The gradient flow (41) is replaced by the flow:

$$\frac{\partial U}{\partial t} + w \cdot \nabla U = 0. \quad (43)$$

The implementation and use of the steepest descent flow on the processor 20 is now described. For example, a discretized algorithm for the AHT flow is presented. The data is naturally provided in a discrete form. The first step is to construct a discrete grid on which $U(x, t)$ will be evaluated. There are many ways for grid generation, and any such grid can be used. Typically, the domain will be a disc, and it is well-known how to construct good grids in discs. The grid on the disc provides the space discretization. In solving the evolution equation (41), the t variable must be discretized. Again, such discretization techniques are well-known, in particular in the context of fluid mechanics and conservation laws.

Denote the discrete time levels by $\{t_i\}$. The initial sampling is used to construct U at time $t_0 = 0$ at the space grid points. For each discrete time step, the vector field v that appears in (41) is found. Once v is known at time step t_i , the value of U at time t_{i+1} is updated using v , the known U at t_i and equation (41). One way to find v at time step t_i is to solve equation (39) using the known U at time t_i , and to use equation (38) to form v . Because the constraint $\nabla \cdot v = 0$ appearing in (38) is very important to satisfy, an alternative method can be used that is based on the stream function notion, such as is very common in fluid mechanics. For this purpose, a stream function $\Psi(x, t)$, such that $\bar{\nabla} \Psi = v$, where the symbol $\bar{\nabla}$ denotes the operator conjugate to the gradient operator, is introduced. This means that the components v_1, v_2 of v are given by:

$$v_1 = -\frac{\partial \Psi}{\partial y}, \quad v_2 = \frac{\partial \Psi}{\partial x}.$$

The advantage of working with Ψ is that v is automatically divergence-free as required. Operating on (38) with $\nabla \times$, and using the fact that no flow exits D_1 , Ψ satisfies the Poisson problem:

$$\Delta \Psi = \nabla \times U \quad x \in D_1, \quad \Psi(x) = 0, \quad x \in \partial D_1. \quad (44)$$

- 5 At each time step t_i the Poisson equation (44) is solved to obtain Ψ at t_i . From Ψ , v is obtained and then (41) is propagated to obtain U at time step t_{i+1} . Equation (44) can be easily solved, for example, by the well-known finite elements method. One proceeds to update U for times t_i until one observes convergence to a minimizer $\bar{U}$. There are several ways to detect convergence. For example, one can check that the updated U for some t_j is almost the same as
- 10 U at time t_{j-1} . Another criterion for convergence is to check whether that v becomes smaller than some predetermined threshold.

Constructing the phase from the optimal mapping $\bar{T}$.

- After the optimal mapping $\bar{T}$ is computed utilizing one of the above-listed algorithms or one substantially similar to it, the phase of the wave can be determined. The optimal mapping $\bar{T}$
- 15 that was computed in the previous examples provide an association between points on two screens. This association is in fact induced by light rays connecting each pair of associated points. Therefore, knowing $\bar{T}$ means that the rays are known. Given a system of rays associated with a propagating wavefront, it is a simple matter to integrate the phase itself, and there are several well-known algorithms for this purpose.

- 20 The method presented hereinabove is based on measuring the intensity on two screens. It is also possible to execute the invention by using three or even more screens. For example, if three screens are used, say screens 1, 2, and 3, then one can first find the ray mapping from

screen 1 to screen 2 and from screen 2 to screen 3, as explained hereinabove. Each such ray mapping provides the rays through screen 2. It might be that the rays obtained for the association of points between screens 1 and 2 will not be exactly the same as the rays obtained between screens 2 and 3. Such difference may be, for example, caused by measurement
5 inaccuracies. One can then average the two values obtained for each ray, and thus reduce the effect of such differences.

The present invention as discussed hereinabove is formulated for monochromatic waves. In practice, the light may be polychromatic. It will be appreciated that the optimization problem ((15), for example) does not depend on the wavelength. Therefore, it can be solved regardless of
10 the wavelength. When determining an actual phase, one must specify a wavenumber k . There are several ways of selecting an effective wavenumber k . For example, in the case where the system is illuminated by a known light source, one can choose as an effective wavenumber k the average wavenumber, weighted with respect to the spectral output of the light source.

While the present invention has been described in detail with reference to certain
15 exemplary embodiments thereof, such are offered by way of non-limiting example of the invention, as other versions are possible. For example, as will be appreciated by one skilled in the art, additional steps might need to be taken for determining the phase of the wave from the intensities. Such additional steps could include, but are not limited to, filtering the intensity measurement to remove noise. It is anticipated that a variety of other modifications and
20 changes will be apparent to those having ordinary skill in the art and that such modifications and changes are intended to be encompassed within the spirit and scope of the invention as defined by the following claims.

Claims

We claim:

1. A device for determining a phase of a wave, the device comprising:
 - (a) a first screen for measuring a first intensity of the wave;
 - 5 (b) a second screen for measuring a second intensity of the wave; and
 - (c) a processor operatively connected to the first and second screens in order to collect the measured first and second intensities, the processor capable of constructing optimal mapping from the first measured intensity and from the second measured intensity, and capable of calculating the phase from the constructed optimal mapping.
- 10 2. The device of claim 1 further comprising a third screen for measuring a third intensity of the wave.
3. The device of claim 2 wherein the processor is operatively connected to the first, second and third screens, the processor capable of constructing optimal mapping from the first measured intensity, the second measured intensity, and the third measured intensity, and capable of
- 15 calculating the phase from the constructed optimal mapping.
4. The device of claims 1 wherein the processor utilizes a linear programming optimization to construct the optimal mapping.
5. The device of claim 1 wherein the processor utilizes a steepest descent flow to construct the optimal mapping.
- 20 6. The device of claim 1 wherein the second screen is located in a different plane than the first screen.
7. The device of claim 1 wherein the wave comprises a nonsymmetric wave.
8. The device of claim 1 wherein the wave comprises a paraxial wave.

9. The device of claim 1 wherein the wave comprises a nonparaxial wave.

10. The device of claim 1 wherein the processor comprises a computer.

11. A method for determining a phase of a wave, the method comprising the steps of:

(a) providing a first screen for measuring a first intensity of the wave, a second screen for
5 measuring a second intensity of the wave, and a processor operatively connected to the
first and second screens;

(b) measuring the first intensity of the wave with the first screen;

(c) measuring the second intensity of the wave with the second screen;

(d) collecting the measured first and second intensities and constructing optimal mapping
10 with the processor from the first intensity and the second intensity; and

(e) calculating the phase of the wave from the constructed optimal mapping with the
processor.

12. The method of claim 11 wherein the providing step further provides a third screen for
measuring a third intensity of the wave.

15 13. The method of claim 12 further comprising the step of measuring the third intensity of the
wave with the third screen.

14. The method of claim 13 wherein the constructing optimal mapping step constructs the
optimal map with the processor from the first, second and third intensities.

15. The method of claim 11 wherein the constructing optimal mapping step comprises the step of
20 utilizing a linear programming optimization in the processor based on the first intensity and the
second intensity to calculate the optimal mapping.

16. The method of claim 11 wherein the constructing optimal mapping step comprises the step of utilizing a steepest descent flow in the processor based on the first intensity and the second intensity to calculate the optimal mapping.

17. The method of claim 11 wherein the method determines the phase of a nonsymmetric wave.

5 18. The method of claim 11 wherein the method determines the phase of a paraxial wave.

19. The method of claim 11 wherein the method determines the phase of a nonparaxial wave.

20. The method of claim 11 where said step of constructing optimal mapping includes the steps of:

(a) using the measured first and second intensities to construct a family of constrained ray mappings and to define a weighted least action functional; and

10

(b) optimizing said weighted least action functional among all mappings in said family of constrained ray mappings.

21. A device for determining a phase of a wave, the device comprising:

(a) at least two detection screens for measuring at least a first intensity and a second intensity at different locations; and

15

(b) a processor operatively connected to the at least two screens in order to collect the measured first and second intensities, the processor capable of constructing optimal mapping from the first measured intensity and from the second measured intensity, and capable of calculating the phase from the constructed optimal mapping.

20 22. The device of claim 21 wherein the at least two detection screens are located in different planes.

23. The device of claim 21 wherein the wave comprises a nonsymmetric wave.

24. The device of claim 21 wherein the wave comprises a paraxial wave.

25. The device of claim 22 wherein the wave comprises a nonparaxial wave.
26. The device of claim 22 wherein the processor comprises a computer.

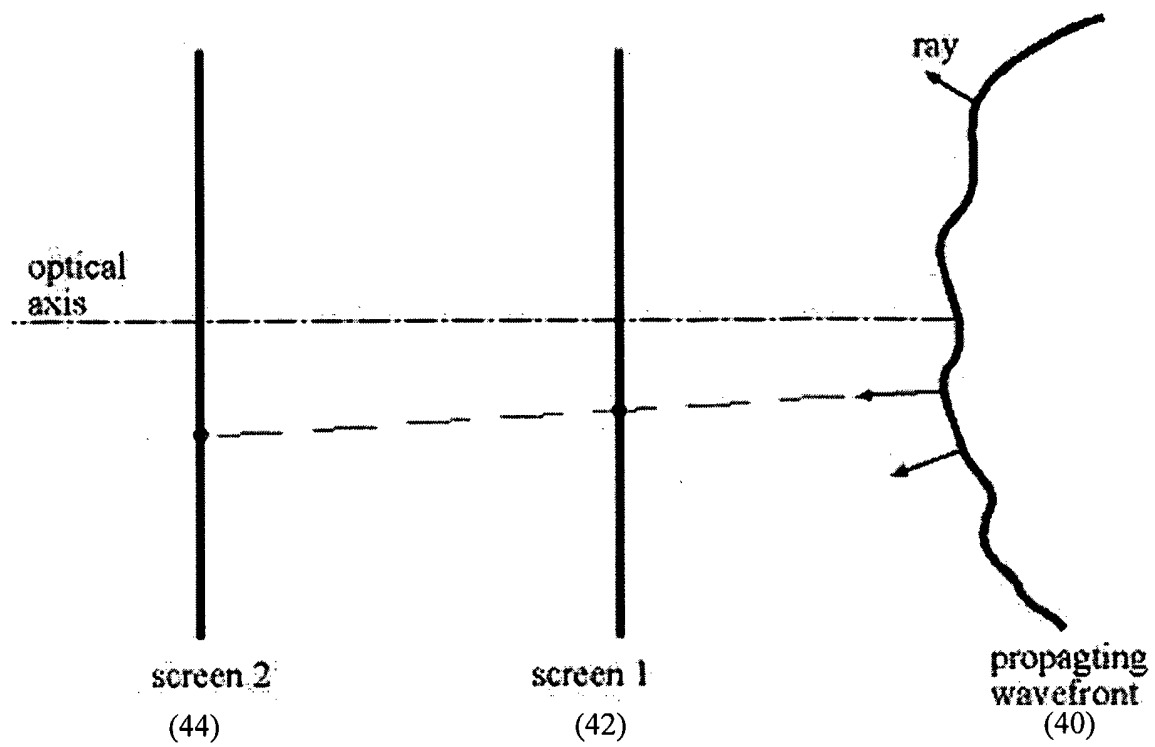


Fig. 1

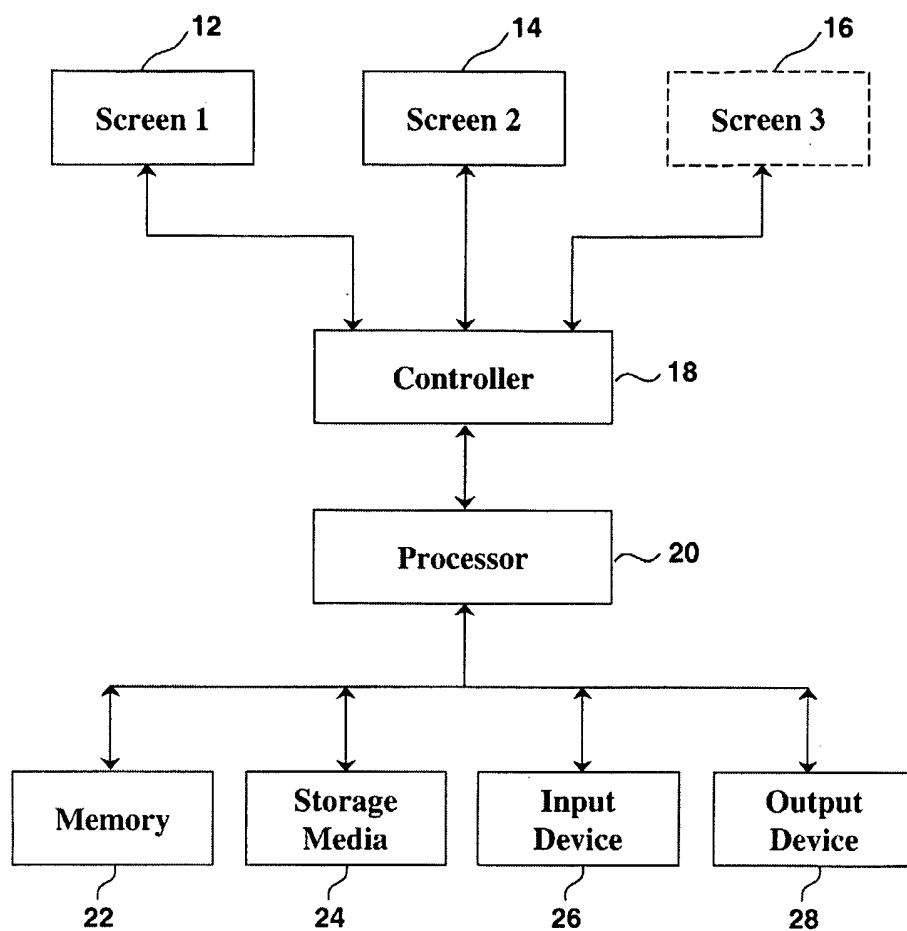


FIG. 2