

ПРИМЕНЕНИЕ МЕТОДА ГРАДИЕНТНОГО СПУСКА ДЛЯ ОПТИМИЗАЦИИ ГОЛОГРАФИЧЕСКИХ ИЗОБРАЖЕНИЙ

Основными задачами голографической литографии являются задачи синтеза голограммы, восстановления изображения и оптимизации качества изображения. Для решения этих задач требуются высокие вычислительные мощности. Все расчёты в данной работе проведены с помощью суперкомпьютеров МИИТ Т-4700 и МВС-100К МСЦ РАН.

METHOD OF STEEPEST DESCENT FOR HOLOGRAPHIC IMAGE OPTIMIZATION / V.V. Chernik (A.J. Ishlinskiy Institute for Problems in Mechanics of the Russian Academy of Sciences, Vernadskovo avenue 101, structure 1, Moscow 119526, Russia, www.ipmnet.ru). The main problems in holographic lithography are hologram synthesis, image modeling and image quality optimization. High computing power is required for solving these problems. Some results in imaging and optimization are shown with all calculations proceeded on МИИТ Т-4700 and МВС-100К.

1. Введение

В работе представлена задача улучшения голографического изображения, полученного при компьютерном моделировании голографических методов создания световых изображений, предложенных и разработанных группой авторов [1], [2]. В указанных работах используется голограмма, рассчитанная по методу Габора [3]. Этот метод позволяет получить так называемую серую голограмму, с помощью которой восстанавливается изображение, которое обычно повторяет заданное с некоторыми отличиями. Дефекты изображения обычно проявляются в виде неравномерной яркости разных элементов, колебаний яркости в пределах одного элемента, недостаточно высокого контраста. Задачей оптимизации является минимизация этих отличий и дефектов изображения за счёт вариаций голограммы. Для определения качества изображения вводятся функционалы оценки отличия изображения от заданного и функционал оптического контраста. Итоговый показатель качества изображения является линейной комбинацией этих функционалов. Голограмма Габора, представленная в модели с помощью вещественнозначной функции на двумерной сетке, используется в качестве первого приближения.

2. Постановка задачи

Исходная задача состоит в создании в плоскости «объекта» O (см. рис. 1) заданного электромагнитного поля, определяемого комплекснозначной функцией $f(x, y)$. Такое электромагнитное поле должно быть создано с помощью облучения голографической пластинки Γ , параллельной плоскости объекта и находящейся на расстоянии

Z от нее, заданной волной, например, сходящейся сферической или плоской.

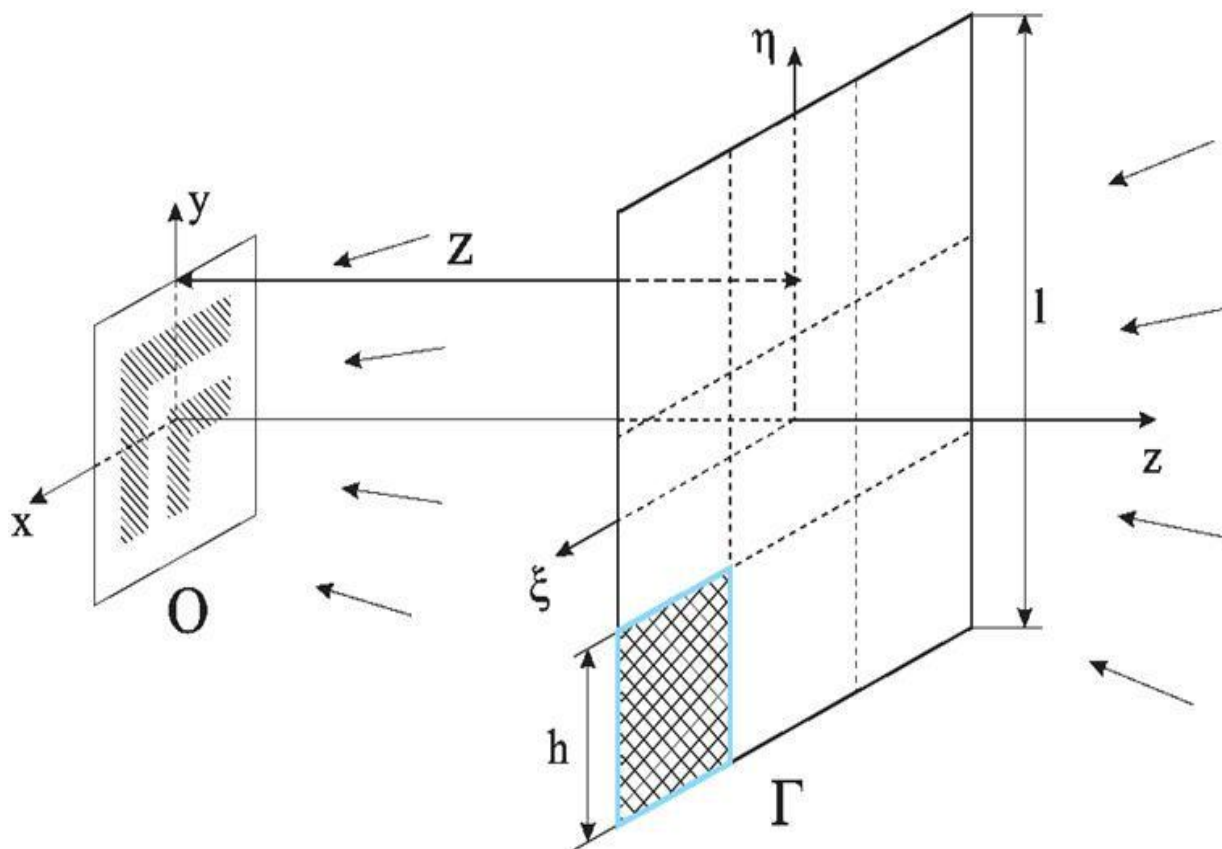


рис. 1

Задача получения комплексной голограммы. Найти такую комплекснозначную функцию $u(\xi, \eta)$, чтобы

$$\iint_{\Gamma} K(x - \xi, y - \eta, Z) u(\xi, \eta) d\xi d\eta = w(x, y),$$

где $w(x, y)$ - заданная на объектной плоскости комплекснозначная функция,

$$K(x, y, z) = -\frac{\partial}{\partial z} \left(\frac{e^{ikr}}{r} \right), \quad r = \sqrt{x^2 + y^2 + z^2}, \quad k = \frac{2\pi}{\lambda}, \quad \lambda - \text{длина волны.}$$

Утверждение. Такая задача может иметь точное решение только в случае, когда $w(x, y)$ имеет финитный спектр.

Если рассматривать заданное поле на объекте O как источник излучения, то комплексную голограмму можно получить обращением поля источника, дошедшего до плоскости Γ . Однако, даже в случае бесконечной голограммы таким образом нельзя точно получить любое изображение.

Дополнительным ограничением является отсутствие технической возможности произвольного влияния на фазу подсвечивающей волны. Можно считать при этом, что фильтр, меняющий амплитуду излучения произвольным образом в любом месте голограммной пластины, реализуем, например, с помощью изменения плотности постановки

одинаковых отверстий малого размера. Поэтому, следует сформулировать задачу получения «серой голограммы».

Задача получения серой голограммы. Найти такую вещественнозначную функцию $U(\xi, \eta)$, чтобы

$$\iint_{\Gamma} K(x - \xi, y - \eta, Z) I(\xi, \eta) U(\xi, \eta) d\xi d\eta = w(x, y),$$

где $I(\xi, \eta)$ - поле волны подсветки.

Для вычислительного эксперимента голограмма разбивается прямоугольной сеткой на ячейки малого размера. Тогда искомая функция принимает вид

$$U(\xi, \eta) = \sum_i \alpha_i \psi_i(\xi, \eta),$$

где $\alpha_i \in R$, $\psi_i(\xi, \eta)$ - функция-индикатор i -й ячейки, равная нулю вне ячейки и единице внутри.

Интеграл в таком случае разбивается на сумму интегралов по ячейкам, для каждого из которых можно применить формулу приближения дальней зоны [4]. Тогда

$$\iint_{\Gamma} K(x - \xi, y - \eta, Z) I(\xi, \eta) U(\xi, \eta) d\xi d\eta \approx \sum_i \alpha_i D_i(x, y).$$

$D_i(x, y)$ можно понимать как поле волны подсветки, прошедшей дифракцию на i -й прямоугольной ячейке. Например, в случае сферической волны подсветки, фокус которой находится в плоскости объекта,

$$D_i(x, y) = \frac{2ie^{ik(R+R_0)}}{\pi R^2} \frac{\sin\left(k\sigma\left(\frac{x-\xi_i}{R} + \frac{x_0-\xi_i}{R_0}\right)\right) \sin\left(k\sigma\left(\frac{y-\eta_i}{R} + \frac{y_0-\eta_i}{R_0}\right)\right)}{k\left(\frac{x-\xi_i}{R} + \frac{x_0-\xi_i}{R_0}\right)\left(\frac{y-\eta_i}{R} + \frac{y_0-\eta_i}{R_0}\right)},$$

где $R = \sqrt{(x - \xi_i)^2 + (y - \eta_i)^2 + Z^2}$, $R_0 = \sqrt{(x_0 - \xi_i)^2 + (y_0 - \eta_i)^2 + Z^2}$, (x_0, y_0) - координаты фокуса волны подсветки на объектной плоскости, (ξ_i, η_i) - координаты i -й ячейки на плоскости голограммы, σ - размер ячейки на голограмме.

Условиями применения приближения дальней зоны являются

$$\frac{\sigma}{Z} \ll 1 \text{ и } \frac{k\sigma^2}{Z} \ll 1.$$

Ясно, что точное решение задачи получения серой голограммы возможно только для узкого класса функций $w(x, y)$. Так же обычно на практике важна задача получения определенного распределения интенсивности излучения на объектной плоскости. А фаза волны, приходящей на объектную плоскость не важна. Отсюда формулируется задача оптимизации.

Задача оптимизации 1. Найти такой набор вещественных чисел $\{\alpha_i\}$, чтобы

$$\iint_0 \left(\left| w(x, y) \right| - \left| \sum_i \alpha_i D_i(x, y) \right| \right)^2 dx dy \rightarrow \min$$

Обратим дальнейшее внимание на следующие практические соображения. Распределение интенсивности поля, создающего изображение, интересует нас только с точки зрения своей формы, а не мощности. То есть, если поле $\sum_i \alpha_i D_i(x, y)$ совпадает с заданным $w(x, y)$ по интенсивности с точностью до умножения на величину, постоянную на всей объектной плоскости, то решение получено. Поэтому, окончательно задачу оптимизации можно сформулировать так.

Задача оптимизации 2. Найти такой набор вещественных чисел $\{\alpha_i\}$, чтобы

$$\rho(w, f) = \iint_0 \left(|w(x, y)| - |f(x, y)| \right)^2 dx dy \rightarrow \min$$

$$\text{где } f(x, y) = \frac{\left| \sum_i \alpha_i D_i(x, y) \right|}{\max_{(a,b) \in O} \left| \sum_i \alpha_i D_i(a, b) \right|}.$$

При численном моделировании объектная плоскость разбивается сеткой, а интеграл, определяющий величину отклонения изображения от заданного, приближается соответствующей суммой. Шаг сетки на изображении определяется разрешением изображения.

3. Получение первого приближения

Для применения итерационных методов таких, как метод градиентного спуска важнейшей задачей является получение первого приближения. От того, насколько удачно выбрано первое приближение, зависит итоговое качество изображения и время на его достижение. Начальная серая голограмма в данной работе синтезируется методом, похожим на метод Габора. Отличия заключаются в выборе внеосевой оптической схемы и расходящейся сферической опорной волны. Основная идея метода в том, что в качестве начальной серой голограммы используется распределение интенсивности поля, являющегося результатом интерференции поля от желаемого изображения, подсвеченного плоской волной, и сферической волны, фокус которой находится в плоскости изображения недалеко от самого изображения.

Таким образом, задача синтеза голограммы, которая будет использована в качестве первого приближения, в самой её трудоемкой части сводится к расчёту поля от желаемого изображения, подсвеченного плоской волной, по скалярным формулам Кирхгофа с применением приближения дальней зоны. Поскольку волна плоская, то интеграл Кирхгофа является сверткой функции изображения и $K(x, y, z) = -\frac{\partial}{\partial z} \left(\frac{e^{ikr}}{r} \right)$, поэтому возможно применение быстрого преобразования Фурье.

4. Выбор размера шага сетки на голограмме

От выбора шага сетки на голограмме зависит время синтеза и оптимизации, которое обратно пропорционально квадрату размера ячейки. Однако функция серости голограммы имеет ярко выраженный осциллирующий характер. Поэтому, выбирая слишком большой шаг сетки на голограмме, можно потерять большую часть информации.

На рис.2 изображен характерный фрагмент серой голограммы.

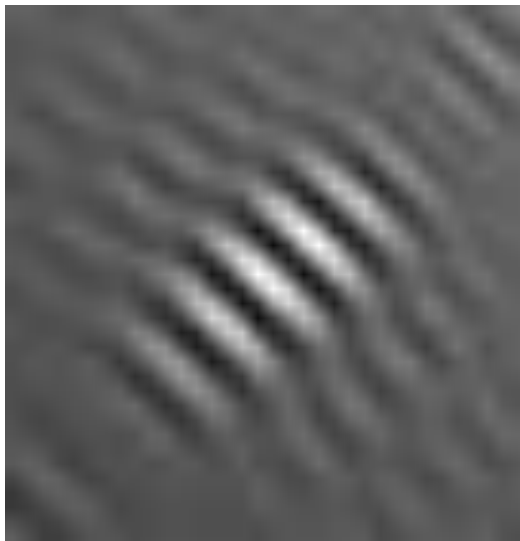


рис.2

Поэтому, для выбора размера шага сетки важно получить оценку характерного периода изменения функции серости.

Утверждение 1. Минимальный период изменения функции серости

$$T \approx \frac{\lambda Z}{S},$$

где λ - длина световой волны используемого излучения, S - расстояние от фокуса опорной сферической волны до наиболее удаленной точки изображения.

Численные эксперименты показали, что достаточно выбирать шаг сетки $\sigma = \frac{\lambda Z}{4S} \approx \frac{T}{4}$.

На рис.3 и рис.4 соответственно приведены исходное идеальное изображение и изображение, восстановленное с голограммы, синтезированной описанным в пункте 3 методом.

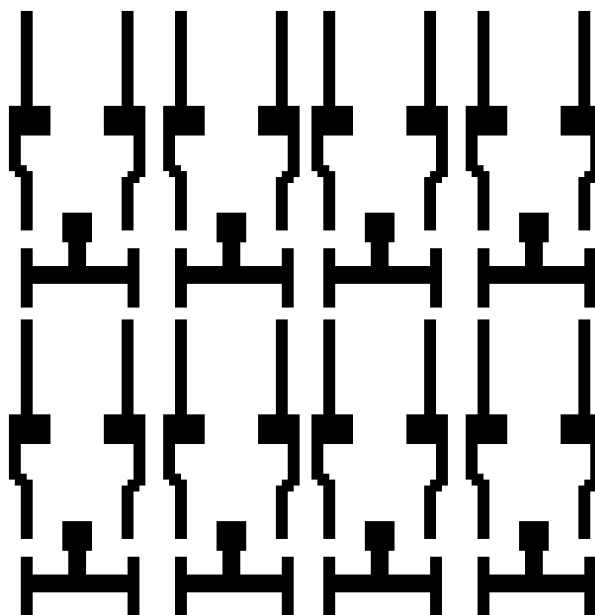


рис.3

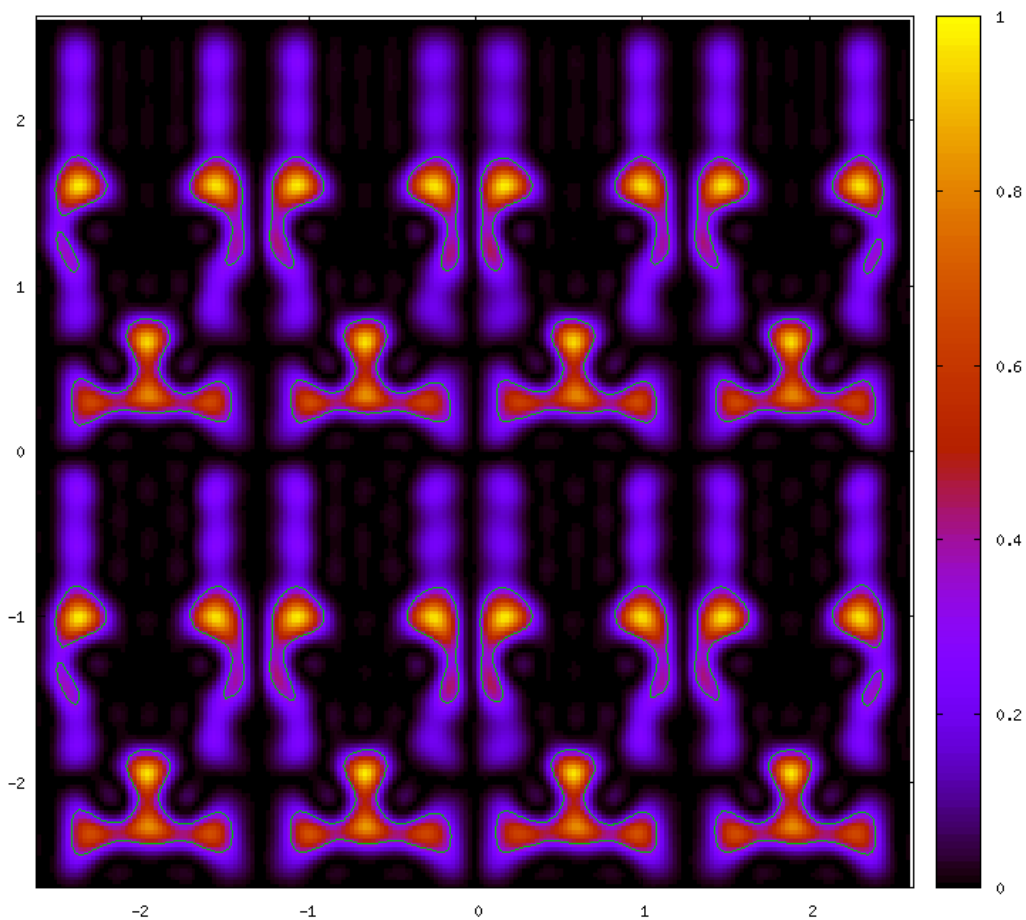


рис.4

На этом восстановленном изображении отчетливо видны неравномерность освещения элементов и эффект Гиббса, выражающийся в волнообразных контурах. Следует отметить, что толщина линий на изображении равна половине длины волны освещения.

Для получения такого изображения следует выбирать расстояние Z между плоскостями изображения и голограммы (см. рис.1) таким, чтобы выполнялись условия применения формулы приближения дальней зоны, а размер голограммы $l = Z$.

Утверждение 2. Количество ячеек на голограмме определяется

$$N = \left(\frac{l}{\sigma}\right)^2 = \left(\frac{4S}{\lambda}\right)^2,$$

где λ - длина световой волны используемого излучения, S - расстояние от фокуса опорной сферической волны до наиболее удаленной точки изображения.

Для получения более сложных изображений, например колец толщиной в половину длины волны, может потребоваться голограмма в два раза больше. Количество ячеек на изображении так же пропорционально $\left(\frac{S}{\lambda}\right)^2$, так как разрешение на изображении имеет смысл выражать в долях длины волны освещения. Поэтому, сложность задачи пропорциональна $\left(\frac{S}{\lambda}\right)^4$.

5. Метод градиентного спуска

Обозначим функцию из задачи оптимизации 2

$$\rho(w, f) = F(\bar{\alpha}),$$

где $\bar{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_N)$ - коэффициенты пропуска ячеек на голограмме. То есть функция отклонения восстановленного изображения от идеального зависит от N вещественных аргументов. Так же пусть $\bar{\alpha}_0$ - первое приближение. Тогда

$$\bar{\alpha}_{t+1} = \bar{\alpha}_t - \beta_t \overline{\text{grad}F}.$$

i -я координата градиента рассчитывается как $F(\dots, \alpha_i + \varepsilon, \dots) - F(\dots, \alpha_i, \dots)$, коэффициент β_t зависит от результатов предыдущих шагов оптимизации.

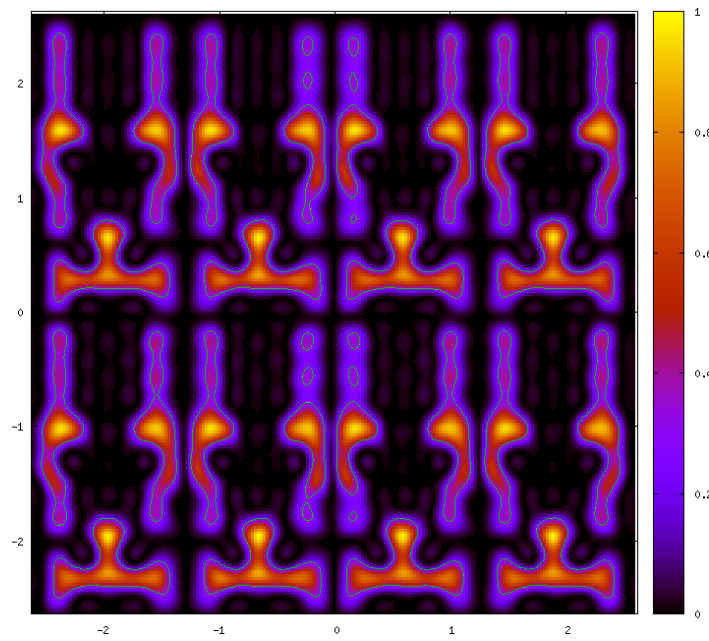
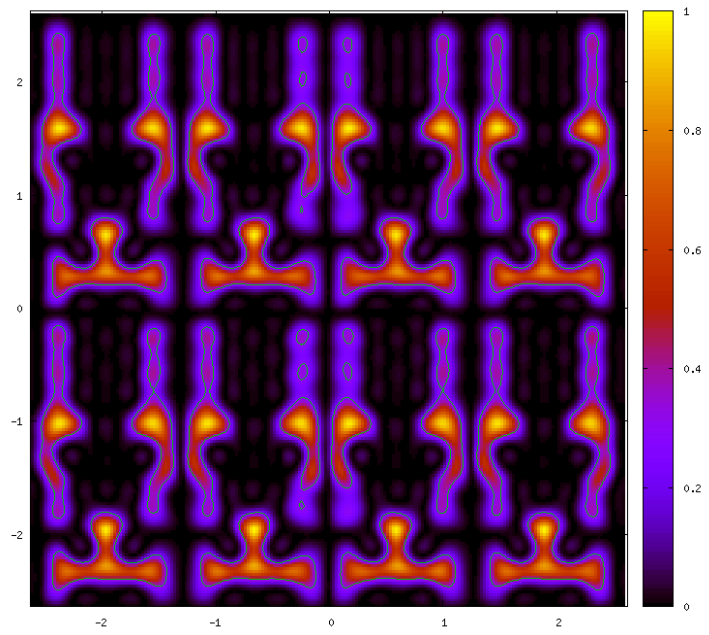
После каждой итерации запускается процесс восстановления изображения и счет градиента. Время на счет градиента примерно равно времени одного восстановления, так как каждая координата градиента рассчитывается добавлением к комплексному полю, приходящему на объект, функции $D_i(x, y)$, умноженной на ε .

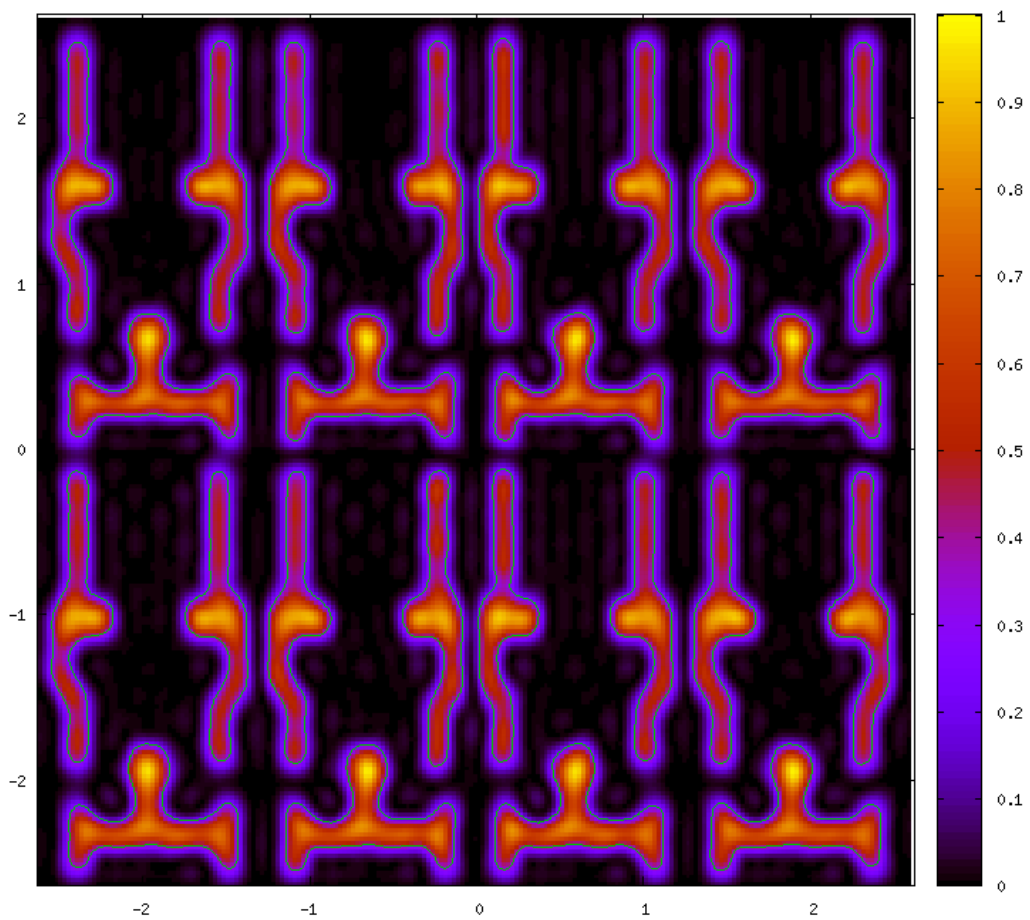
При расчете голограмма разбивается на блоки, как показано на рис.1. Поле от одного блока голограммы рассчитывается одним ядром.

6. Некоторые результаты

Одна итерация для изображения на рис.3 на 128 ядрах процессора AMD Opteron™ 2356 (Barcelona) (МИИТ Т-4700) занимала около 4 секунд. Размер изображения около 26,5 мкм². Длина волны 196 нм.

На следующих рисунках показаны результаты восстановления после 10, 20 и 100 итераций.

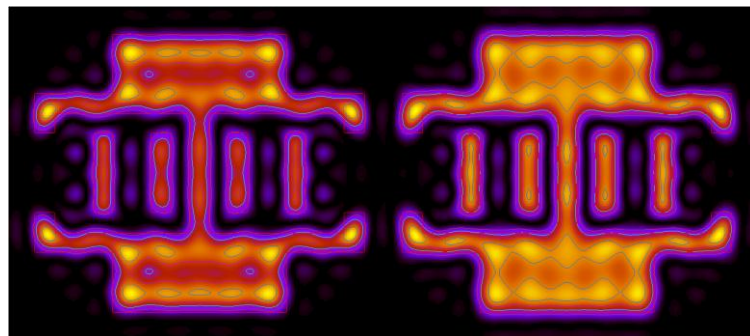




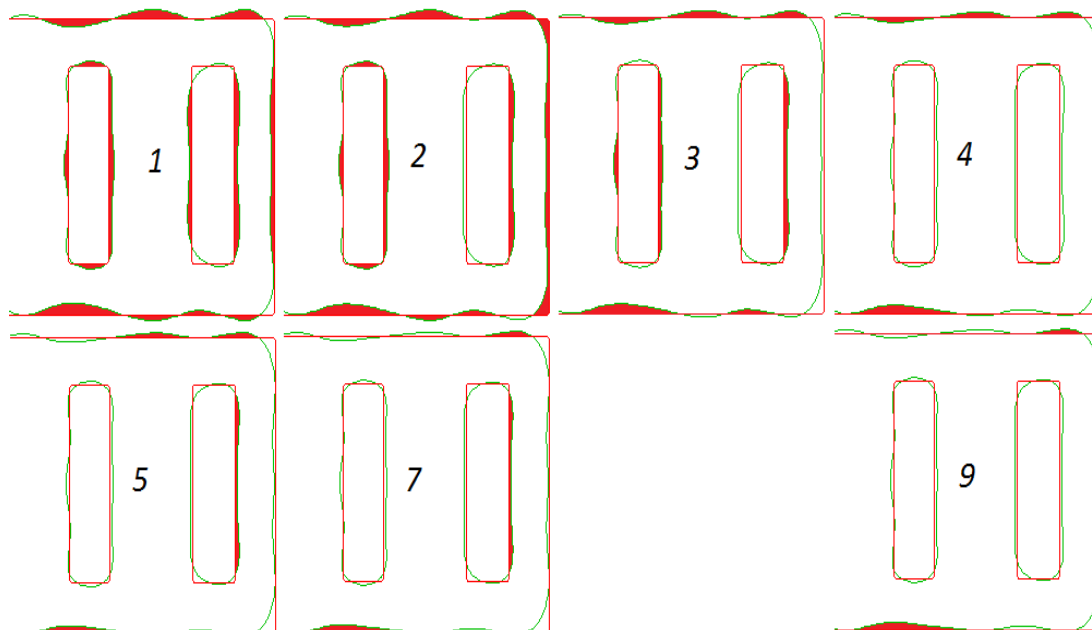
Норма ошибки после 100 итераций снижена почти в два раза.

Важным показателем качества изображения является наличие контура на одном уровне освещения, попадающего в некоторый коридор контура идеального изображения.

Изображение объекта до и после 9 итераций:



На следующем рисунке показана эволюция контура фрагмента изображения.



СПИСОК ЛИТЕРАТУРЫ

1. Борисов М.В., Боровиков В.А., Гавриков А.А., Князьков Д.Ю., Раховский В.И., Челюбеев Д.А., Шамаев А.С. Методы создания и коррекции качества голографических изображений геометрических объектов с элементами субволновых размеров // Доклады Академии Наук. 2010. Т. 434, № 3. С. 332–336.
2. Борисов М.В., Гавриков А.А., Князьков Д.Ю., Раховский В.И., Челюбеев Д.А., Шамаев А.С. Способ изготовления голографических изображений рисунка // Патент РФ 2396584 С1. — 2010.
3. Gabor D. A. A New Microscopic Principle // Nature, 1948, V161, p. 777-778
4. Басс Ф.Г., Фукс И.М. Рассеяние волн на статистически неровной поверхности. М: Наука, 1972.
5. Н. Н. Моисеев, Ю. П. Иванюков, Е. М. Столярова. Методы оптимизации. М.: Наука, 1978.
6. Черноушко Ф.Л., Баничук Н.В. Вариационные задачи механики и управления. - Численные методы. М.: Наука, 1973.
7. Князьков Д.Ю. Задачи расчёта и оптимизации электромагнитных полей голографической литографии // Доклады пятой международной конференции «Параллельные вычисления и задачи управления». 2010. С. 425-432.