

ОЦЕНКА ГЕОМЕТРИЧЕСКИХ ИСКАЖЕНИЙ КОНТУРОВ ИЗОБРАЖЕНИЙ ГУБ В СИСТЕМАХ ВИЗУАЛЬНОГО ВВОДА ИНФОРМАЦИИ

Р. Г. Хафизов^а, доктор техн. наук, профессор

Т. В. Яранцева^а, аспирант

^аПоволжский государственный технологический университет, Йошкар-Ола, РФ

Постановка проблемы: чтение по губам является одним из способов извлечения данных о речи из видеоинформации. Существующие системы распознавания речи по изображениям губ достаточно эффективно работают при фронтальном расположении говорящего к видеокамере, т. е. когда на изображениях наблюдаются только линейные искажения, не приводящие к изменению формы губ. При расположении говорящего к видеокамере под углом наблюдаются нелинейные искажения. Форма губ при этом изменяется. Вопросы компенсации нелинейных искажений практически не рассмотрены. **Цель:** разработка алгоритма оценки геометрических искажений контуров губ говорящего, который позволит усовершенствовать существующие системы ввода речевой информации из видеоданных. **Результаты:** предложена методика оценки геометрических искажений контуров губ в системах визуального ввода информации. Оценка параметра геометрического искажения производится по результатам вычисления нормированного скалярного произведения наблюдаемого контура изображения губ говорящего с преобразованными контурами при варьировании параметра искажения. **Практическая значимость:** предложенный алгоритм оценки вида и параметра искажения контура позволяет повысить эффективность распознавания искаженных контуров изображений губ говорящего в системах визуального ввода речевой информации из видеоданных.

Ключевые слова — контур, математическая модель, геометрические искажения, распознавание речи.

Введение

Система визуального ввода информации, являющаяся важным направлением усовершенствования человеко-машинных интерфейсов, позволяет упростить и автоматизировать речевой ввод [1]. При этом одним из основных способов извлечения данных о речи из видеоинформации является чтение по губам, основанное на распознавании контуров губ [2, 3]. Контурный анализ подразумевает такой способ обработки изображений, при котором обработка подвергается лишь контур изображения, в то время как обработка каждой точки изображения не требуется. Такой подход позволяет не рассматривать внутренние точки изображения и тем самым значительно сократить объем обрабатываемой информации [4].

В настоящее время предложены различные подходы к выделению изображения лица человека и контура его губ. Так, в работе [2] предложен способ формирования сообщений на основе детектирования движения губ. С помощью метода активного контура строится геометрическая модель лица человека. Для устранения неинвариантности дескрипторов изображений рта авторами предложен алгоритм предварительной обработки изображений, состоящий в масштабировании исходных размеров изображения с переходом к стандартной величине по ширине или высоте. В работе [3] для решения задачи распознавания контуров губ в видеопотоке используются цветные изображения, и цвет выбран в качестве

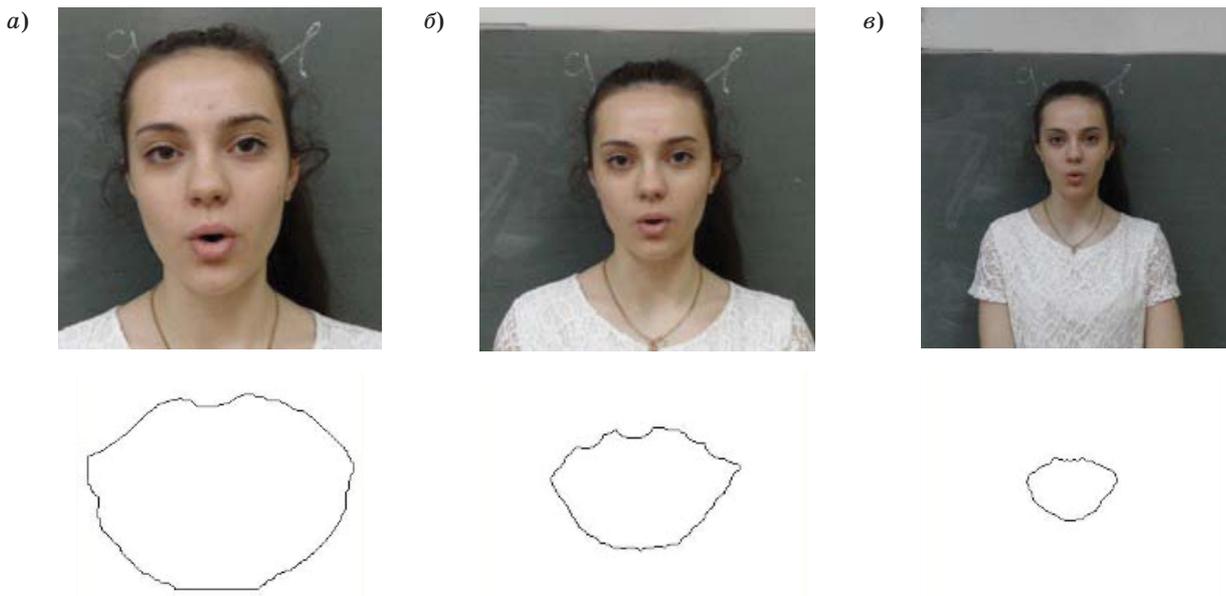
главного источника информации при выделении контура губ. В работе [5] предложено рассматривать контуры изображений губ как совокупность возможных реализаций некоторой случайной функции и для их анализа использовать теорию случайного контура [6, 7].

Существующие системы распознавания речи по изображениям губ достаточно эффективно работают при фронтальном расположении говорящего к видеокамере, т. е. когда на изображениях наблюдаются только линейные искажения, не приводящие к изменению формы губ (изменение масштаба и угла поворота в плоскости кадра). Вопросы же компенсации нелинейных искажений, при которых форма губ изменяется, практически не рассмотрены.

Математическая модель контура изображения губ

Контур задается как комплексная случайная функция неслучайного аргумента l : $X(l) = \text{Re}X(l) + i\text{Im}X(l)$. Случайный контур $X(l)$ рассматривается как совокупность его возможных реализаций $\chi_1(l), \chi_2(l), \dots, \chi_n(l)$, где n — количество реализаций [6]. Контур-реализация $\chi_k(l)$ случайного контура $X(l)$ представляет собой периодическую последовательность элементарных векторов, заданных в виде комплексных чисел: $\chi_k(l) = \text{Re}\chi_k(l) + i\text{Im}\chi_k(l)$.

Для случайного контура $X(l)$ введены понятия математического ожидания и дисперсии.



■ **Рис. 1.** Пример линейного искажения в виде изменения масштаба: $a - \mu > 1$; $b - \mu = 1$; $c - \mu < 1$
 ■ **Fig. 1.** The example of linear distortion in the form of change of scale: $a - \mu > 1$; $b - \mu = 1$; $c - \mu < 1$

При этом математическое ожидание случайного контура $X(l)$ графически истолковывается как «усредненный контур», около которого расположены другие контуры-реализации $\chi_1(l), \chi_2(l), \dots, \chi_n(l)$. Дисперсия характеризует степень рассеяния контуров-реализаций $\chi_1(l), \chi_2(l), \dots, \chi_n(l)$ вокруг математического ожидания случайного контура $X(l)$ [6]. Также заданы корреляционная $K_X(l_1, l_2)$ и взаимная корреляционная $K_{XY}(l_1, l_2)$ функции случайного контура $X(l)$.

Для формирования модели контуров изображений губ говорящего необходимо зафиксировать изображения губ при произнесении различных слов. Выделив контуры изображений губ, получим выборки, содержащие контуры-реализации изображений губ при произнесении разных звуков.

В работе [4] показано, что применение комплекснозначного описания кода контура позволяет получить инвариантность величины модуля нормированного скалярного произведения к линейным искажениям контура типа масштабирование и поворот. При этом под поворотом контура на угол φ понимается линейное преобразование, в результате которого на этот угол поворачивается каждый элементарный вектор контура. Аналогично изменение масштаба контура в μ раз заключается в растяжении каждого элемента контура в μ раз. На рис. 1, $a-c$ приведены изображения лица и контуры губ при изменении расстояния до камеры.

Значение нормированного скалярного произведения между любой парой контуров, пред-

ставленных на рис. 1, постоянно и близко к единице. Таким образом, линейное искажение типа масштабирование практически не влияет на качество распознавания контуров губ говорящего в системах визуального ввода информации.

Оценка геометрического искажения контура изображения губ

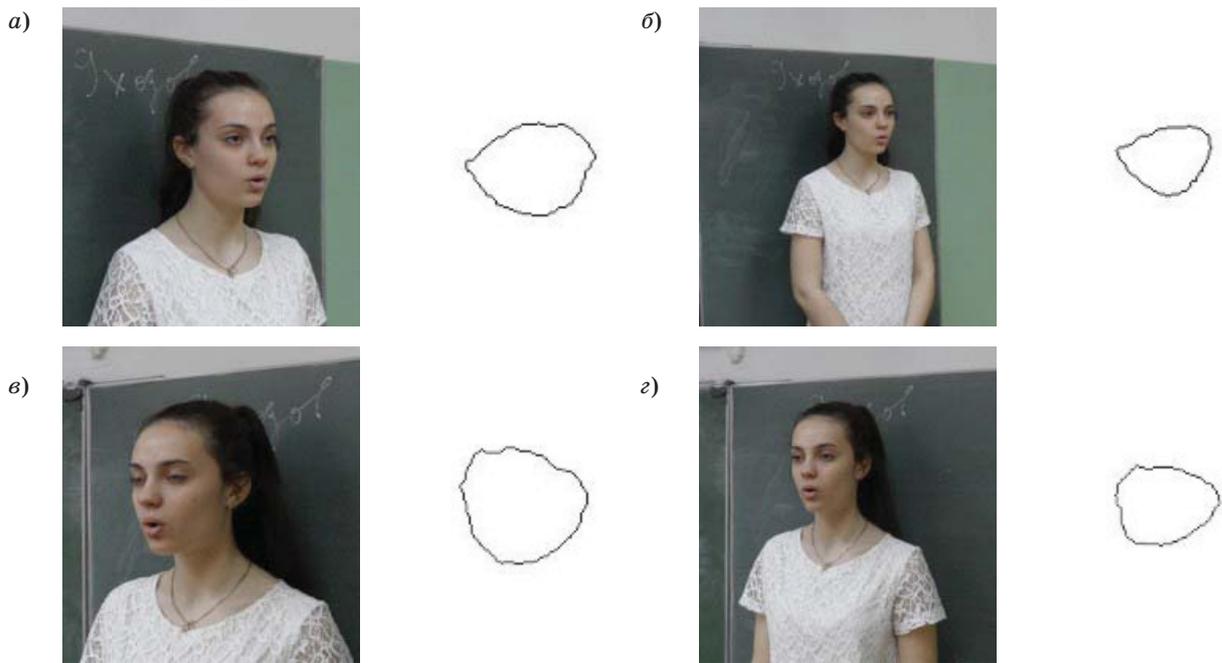
Геометрическое искажение контура изображения губ можно наблюдать, например, при расположении говорящего к видеокамере под некоторым углом (рис. 2, $a-z$). Характер искажения в этом примере относится к так называемым трапецидальным искажениям.

В работе [8] представлена модель контура, подвергнутого трапецидальному искажению:

$$v(n) = \gamma(n) + iA \exp\left(i \frac{2\pi}{s} 2n\right), \quad n = 0, 1, \dots, s-1,$$

где $v(n)$ — код контура искаженного изображения; $\gamma(n)$ — код контура исходного изображения; A — коэффициент искажения; s — длина контура. Показано, что при трапецидальном искажении к спектру исходного контура добавляется одна гармоника, уровень которой определяется параметром A . В зависимости от того, является эта гармоника чисто вещественной или мнимой, определяется характер трапецидального искажения.

Применение предложенных [8] моделей контуров искаженных изображений позволяет моде-



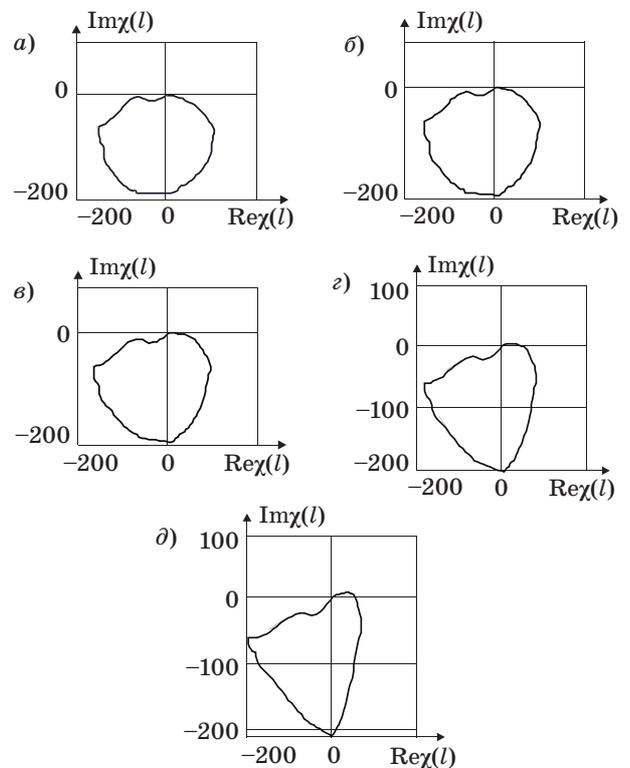
■ **Рис. 2.** Примеры трапецидальных искажений, возникающих при расположении говорящего к видеокамере: *a, б* — справа; *в, з* — слева

■ **Fig. 2.** The examples of the trapezoidal distortions arising at an arrangement of the speaker to the video camera: *a, б* — on the right; *в, з* — at the left

лизовать эталонные контуры для сравнения их с наблюдаемыми и производить оценку параметров и вида искажений. На рис. 3, *a–д* представлен пример преобразованного контура изображения губ при изменении параметра A для трапецидального искажения.

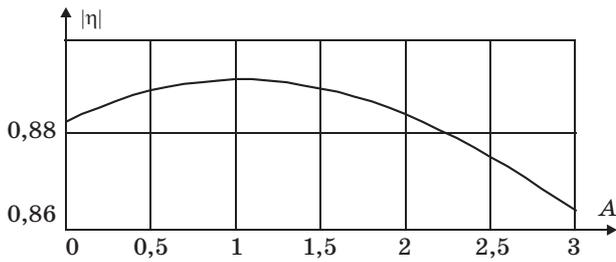
Оценка параметра A трапецидального искажения может быть найдена по результатам вычисления нормированного скалярного произведения $\eta = (N, \Gamma_A)$ наблюдаемого контура N изображения губ говорящего (см. рис. 2) с искаженными контурами Γ_A при изменении параметра искажения A (см. рис. 3). Наблюдаемый контур N представляет собой последовательность элементарных векторов $v(n)$, заданных в виде комплексных чисел $N = \{v(n)\}$, где $v(n) = \text{Re}v(n) + i\text{Im}v(n)$, $n = 0, 1, \dots, s - 1$. Искаженный контур Γ_A также представляет собой последовательность комплексных чисел $\gamma_A(n) = \text{Re}\gamma_A(n) + i\text{Im}\gamma_A(n)$, $n = 0, 1, \dots, s - 1$. При этом совокупность контуров Γ_A формируется из эталонного контура Γ с помощью преобразования «трапецидальное искажение» при изменении параметра искажения A . Нормированное скалярное произведение вычисляется следующим образом:

$$\eta = (N, \Gamma_A) = \frac{\sum_{n=0}^{s-1} v(n)\gamma_A^*(n)}{\|N\| \|\Gamma_A\|}.$$

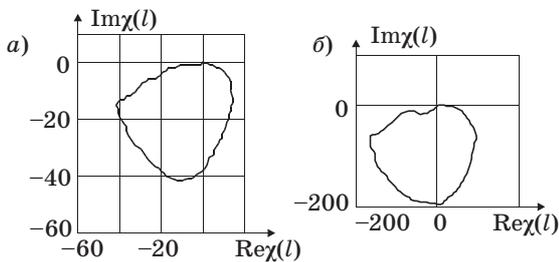


■ **Рис. 3.** Вид искаженного контура изображения губ при изменении параметра A : *a* — 0; *б* — 0,5; *в* — 1; *з* — 2; *д* — 3

■ **Fig. 3.** The type of the distorted contour of the image of lips at change of parameter A : *a* — 0; *б* — 0,5; *в* — 1; *з* — 2; *д* — 3



■ **Рис. 4.** Зависимость модуля нормированного скалярного произведения от параметра A
 ■ **Fig. 4.** The dependence of the module of the normalized scalar product from the parameter A



■ **Рис. 5.** Наблюдаемый контур изображения губ (а) и результат моделирования с оценкой параметра A (б)
 ■ **Fig. 5.** The observed contour of the image of the lips (а) and the result of modeling with an estimate of the parameter A (б)

На рис. 4 представлена зависимость модуля нормированного скалярного произведения наблюдаемого контура N , показанного на рис. 2, б, и искаженного контура изображения губ от параметра A (см. рис. 3).

Литература

1. Калиновский И. А., Спицин В. Г. Обзор и тестирование детекторов фронтальных лиц // Компьютерная оптика. 2016. Т. 40. № 1. С. 99–111. doi:10.18287/2412-6179-2016-40-1-99-111
2. Ковшов Е. Е., Завистовская Т. А. Система обработки движения губ человека для речевого ввода информации // Cloud of Science. 2014. Т. 1. № 2. С. 279–291.
3. Солдатов С. А. Чтение по губам: распознавание контуров губ // Лаборатория компьютерной графики. <http://www.ict.edu.ru/ft/002415/loi2002sss.pdf> (дата обращения: 10.01.2017).
4. Фурман Я. А., Кревецкий А. В., Передреев А. К. и др. Введение в контурный анализ и его приложение к обработке изображений и сигналов / под ред. Я. А. Фурмана. — М.: Физматлит, 2002. — 592 с.

Результаты исследования зависимости модуля нормированного скалярного произведения η от параметра A позволяют произвести оценку параметра искажения A и получить модель наблюдаемого контура изображения губ говорящего (рис. 5, а и б). При этом выбирается то значение параметра искажения A , при котором величина модуля нормированного скалярного произведения η достигает максимума. В примере, представленном на рис. 4, оценка величины A , при которой модуль нормированного скалярного произведения η достигает максимума, равна 1.

Таким образом, применение моделей контуров искаженных изображений позволяет формировать эталонные контуры для сравнения их с наблюдаемыми и производить оценку параметров и вида искажений. Учет этих данных может быть применен для усовершенствования существующих систем ввода речевой информации из видеоданных, основанных на распознавании контуров изображений губ говорящего.

Заключение

В работе для решения задачи распознавания изображений губ говорящего предложен подход на основе методов контурного анализа. Эти методы используют информацию, имеющуюся в контурах изображений, и инвариантны к параметрам линейных преобразований. Показана возможность применения теории случайного контура для реализации системы распознавания речи по контурам изображений губ. Учет моделей контуров искаженных изображений способствует усовершенствованию существующих человеко-машинных интерфейсов ввода речевой информации из видеоданных.

5. Хафизов Р. Г., Яранцева Т. В., Охотников С. А. Система распознавания речи по контурам изображения губ // Проектирование и технология электронных средств. 2016. № 2. С. 13–17.
6. Хафизов Р. Г. Модель контура изображения со случайной формой // Компьютерная оптика. 2014. № 1. С. 127–130.
7. Хафизов Р. Г., Роженцов А. А., Хафизов Д. Г., Охотников С. А. Основы теории обработки непрерывных контуров изображений: монография / под общ. ред. Р. Г. Хафизова. — Йошкар-Ола: ПГТУ, 2015. — 172 с.
8. Хафизов Р. Г., Охотников С. А., Яранцева Т. В. Модели контуров изображений объектов при наличии геометрических искажений // Компьютерная оптика. 2016. Т. 40. № 3. С. 415–420. doi:10.18287/2412-6179-2016-40-3-415-420

UDC 621.391

doi:10.15217/issn1684-8853.2017.4.2

Estimation of Geometrical Distortions of Lip Contours in Visual Input Systems

Khafizov R. G.^a, Dr. Sc., Tech., Professor, HafizovRG@volgatech.netYaranceva T. V.^a, Post-Graduate Student, fiza08@rambler.ru^aVolga State University of Technology, 3, Lenin Sq., 424000, Yoshkar-Ola, Russian Federation

Introduction: Lip reading is a method of extracting speech data from video information. The existing lip-image speech recognition systems are sufficiently powerful when the speaker is located en face in front of the camera, i.e. when the images are observed with only linear distortions, not causing any change in the shape of the lips. When the speaker is located at an angle, nonlinear distortions come out, and the shape of the lips changes. The problems of compensating the nonlinear distortions have almost never been discussed.

Purpose: The goal is to develop an algorithm for estimating the geometric distortions of the lip contours which would make it possible to improve the existing systems of retrieving voice information from video data. **Results:** A technique has been proposed for estimating geometrical distortions of lip contours in visual information input systems. The geometric distortion parameter is estimated by the results of calculating the normalized scalar product of the observed contour of the speaker's lips and the transformed contours, as the distortion parameter is changing. **Practical relevance:** The proposed algorithm for the estimation of the type and parameter of a contour distortion allows you to promote the efficiency of recognizing distorted contours of lip images in systems of visual input of speech information from video data.

Keywords — Contour, Mathematical Model, Geometrical Distortions, Speech Recognition.

References

1. Kalinovskii I. A., Spitsyn V. G. Review and Testing of Frontal Face Detectors. *Komp'iuternaia optika* [Computer Optics], 2016, vol. 40, no. 1, pp. 99–111 (In Russian). doi:10.18287/2412-6179-2016-40-1-99-111
2. Kovshov E. E., Zavistovskaya T. A. Development of Software for Testing Algorithms Design Information Structures. *Cloud of Science*, 2014, vol. 1, no. 2, pp. 279–291 (In Russian).
3. Soldatov S. A. Lip Reading: Lip Contour Recognition. *Laboratoriia komp'iuternoii grafiki*. Available at: <http://www.ict.edu.ru/ft/002415/loi2002sss.pdf> (accessed 10 January 2017) (In Russian).
4. Furman Ya. A., Krevetsky A. V., Predreyev A. K., et al. *Vvedenie v konturnyi analiz i ego prilozhenie k obrabotke izobrazhenii i signalov* [Contour Analysis Introduction and its Image and Signal Processing Application]. Ed. by Ya. A. Furman. Moscow, Fizmatlit Publ., 2002. 592 p. (In Russian).
5. Khafizov R. G., Yaranceva T. V., Okhotnikov S. A. Speech Recognition System for Image Contours Lips. *Proektirovanie i tekhnologiya elektronnykh sredstv* [Design and Technoigy of Electronic Means], 2016, no. 2, pp. 13–17 (In Russian).
6. Khafizov R. G. Model of the Contour of the Image with a Random Shape. *Komp'iuternaia optika* [Computer Optics], 2014, no. 38(1), pp. 127–130 (In Russian).
7. Khafizov R. G., Rozhentsov A. A., Khafizov D. G., Okhotnikov S. A. *Osnovy teorii obrabotki nepreryvnykh konturov izobrazhenii* [Basics of Continuous Image Processing Circuits Theory]. Ed. by R. G. Khafizov. Yoshkar-Ola, Povolzhskii gosudarstvennyi tekhnologicheskii universitet Publ., 2015. 172 p. (In Russian).
8. Khafizov R. G., Okhotnikov S. A., Yaranceva T. V. Models of the Image of Object Contours with Geometrical Distortions. *Komp'iuternaia optika* [Computer Optics], 2016, no. 40(3), pp. 415–420 (In Russian). doi:10.18287/2412-6179-2016-40-3-415-420

УВАЖАЕМЫЕ АВТОРЫ!

Научные базы данных, включая SCOPUS и Web of Science, обрабатывают данные автоматически. С одной стороны, это ускоряет процесс обработки данных, с другой — различия в транслитерации ФИО, неточные данные о месте работы, области научного знания и т. д. приводят к тому, что в базах оказывается несколько авторских страниц для одного и того же человека. В результате для всех по отдельности считаются индексы цитирования, снижая рейтинг ученого.

Для идентификации авторов в сетях Thomson Reuters проводит регистрацию с присвоением уникального индекса (ID) для каждого из авторов научных публикаций.

Процедура получения ID бесплатна и очень проста: входите на страницу <http://www.researcherid.com>, слева под надписью «New to ResearcherID?» нажимаете на синюю кнопку «Join Now It's Free» и заполняете короткую анкету. По указанному электронному адресу получаете сообщение с предложением по ссылке заполнить полную регистрационную форму на ORCID. Получаете ID.