

## Новый инвариантный к освещению дескриптор признаков для распознавания объектов

**Х. Диаз-Эскобар\***, **В.И. Кобер\*\***, **В.Н. Карнаухов\*\***, **Х. Гонзалез-Фрага\*\*\***

*\* Департамент Компьютерных наук, Центр Научных исследований и высшего образования, Энсенада, 22860*

*\*\* Институт проблем передачи информации, Российская академия наук, Москва, 127051, Россия*

*\*\*\* Автономный университет Нижней Калифорнии, Энсенада, 22860*

Поступила в редколлегию 29.01.2018

**Аннотация**—В данной работе предложен новый инвариантный к неравномерному освещению дескриптор для описания признаков на цветных изображениях. Предлагаемый метод построения дескриптора основан на использовании локальной модели энергии, как биологически правдоподобной модели визуальной системы. Алгоритм обнаружения и построения дескриптора использует моногенное масштабируемое пространство и модифицированный алгоритм вычисления гистограммы ориентированных градиентов на основе фазовой конгруэнтности сигналов. Результаты компьютерного моделирования показывают, что предлагаемый дескриптор обеспечивает превосходное обнаружение признаков при неравномерном освещении, шуме и незначительных геометрических искажений по сравнению с известными дескрипторами.

**КЛЮЧЕВЫЕ СЛОВА:** дескриптор, неравномерное освещение, фазовая конгруэнтность, моногенный сигнал, гистограмма ориентированных градиентов.

### ВВЕДЕНИЕ

Несмотря на достигнутый прогресс в области обработки изображений и компьютерного зрения, проблема распознавание изображений по-прежнему остается актуальной [1]. Выделение признаков с помощью локальных дескрипторов и статистические модели – это два подхода, которые обычно используются для решения проблемы [2, 3]. Задача выделения локальных признаков состоит в том, чтобы найти такие объекты на изображении (особые точки, линии, области и т. д.), которые являются локальными, отличительными и повторяемыми характеристиками [4]. Для того, чтобы выделить такие признаки, обычно рассматривается локальная окрестность вокруг каждого особого объекта, которая кодируется в дескриптор. Для поиска соответствия между изображениями наборы дескрипторов двух изображений сравниваются с использованием некоторой метрики, например расстояния Махаланобиса (мера расстояния между векторами, обобщающая понятие евклидова расстояния). Желательно, чтобы дескрипторы были инвариантными к геометрическим искажениям и к неравномерному освещению, а также были робастными к шуму [4]. В реальных приложениях существующие дескрипторы по-прежнему являются недостаточно информативными из-за комплексного ухудшения качества изображения, что снижает эффективность современных методов распознавания, основанных на дескрипторном описании изображений. Отметим, что визуальная система человека способна распознавать объекты с разной освещенностью. Человеческий глаз воспринимает энергию света, отраженную от поверхности объекта (яркость). Яркость преобразуется в нервные импульсы фоторецептором клеток в сетчатке, где информация кодируется и отправляется в первичную визуальную кору (V1) [5]. Психофизические данные свидетельствуют о том, что

визуальная система человека раскладывает визуальную информацию на области границ и линий, используя фазовую информацию. Более того, известно, что различные группы ячеек в V1 извлекают особые признаки изображения, такие как частоту, ориентацию и фазу [6]. В этой работе предлагается фазовый метод вычисления дескриптора, основанный на использовании локальной модели энергии, как биологически правдоподобной модели визуальной системы. Предлагаемый метод выделяет наиболее значительные локальные особенности на изображениях и является инвариантным к неравномерному освещению, геометрическим искажениям и робастным к шуму. Статья организована следующим образом: в разделе 1 описана предлагаемый подход к построению дескрипторов, в разделе 2 представлены полученные результаты компьютерного моделирования, и раздел 3 суммирует наши выводы.

## 1. МОДЕЛЬ МОНОГЕННОГО СИГНАЛА

В работах [7–9] предложена локальная энергетическая модель восприятия границ объектов, линий и теней. Согласно этой модели, визуальная система человека способна определять квадратную и трапециевидную форму волны с использованием информации о фазе, и можно доказать, что максимум энергетической функции происходит в точках с максимальной фазовой конгруэнтностью [10]. Была предложена безразмерная мера фазовой конгруэнтности в каждой точке изображения [11–13], где значение фазовой конгруэнтности – значимость признака в данной точке; единица означает наиболее значимый признак, и ноль – самое низкое значение признака. В работе [14] предложена новая концепция двумерного аналитического сигнала, называемого моногенным сигналом.

Пусть  $H_e \in L^2$  и  $H_o \in L^2$  – базовые операторы модели с равными амплитудами спектра и ортогональными фазами;  $H_o$  – преобразование Гильберта  $H_e$ . Локальная энергетическая функция определяется как

$$E(x) = \sqrt{(H_e(x) * f(x))^2 + (H_o(x) * f(x))^2}, \quad (1)$$

где  $f(x) \in L_2$  – периодический сигнал,  $*$  – оператор свертки.

Локальная энергетическая функция локализует признаки на изображении, но не содержит информации о типе признака. Чтобы определить тип признака, необходимо вычислить функцию аргумента, определенную как

$$\varphi(x) = \arctan \left( \frac{H_e(x) * f(x)}{H_o(x) * f(x)} \right), \quad (2)$$

Периодическая функция  $f(x)$  может быть разложена в ряд Фурье:

$$f(x) = \sum_n A_n \cos(\varphi_n(x)), \quad (3)$$

где  $A_n$  – амплитуда  $n$ -ой компоненты в разложении Фурье, и  $\varphi_n(x) = n\omega x + \phi(x)$  – локальная фазовая функция. Функция фазовой конгруэнтности определяется как [8]

$$PC(x) = \max_{\varphi(x) \in [0, 2\pi]} \frac{\sum_n A_n \cos(\varphi(x) - \bar{\varphi}(x))}{\sum_n A_n}, \quad (4)$$

где  $\bar{\varphi}$  – взвешенное среднее значение локальной фазовой функции по всем Фурье компонентам в данной точке. Так как данное определение чувствительно к шуму, то робастный вариант фазовой конгруэнтности можно определить как [11]

$$PC(x) = \frac{W(x)[E(x) - T]}{\sum_n A_n(x) + \varepsilon}, \quad (5)$$

где  $W(x)$  – весовая функция,  $E(x)$  – энергия сигнала,  $T$  – порог, зависящий от шума,  $\varepsilon$  – малая константа для исключения деления на ноль.

Моногенный сигнал [15] определяется как комбинация двумерного сигнала и его преобразование Риеса первого порядка. Пусть  $R_x; R_y$  – передаточные функции преобразования Риеса первого порядка в частотной области:

$$R_x(u, v) = i \frac{u}{\sqrt{u^2 + v^2}} = F_r \left( \frac{x}{2\pi(x^2 + y^2)^{\frac{3}{2}}} \right), \quad (6)$$

$$R_y(u, v) = i \frac{v}{\sqrt{u^2 + v^2}} = F_r \left( \frac{y}{2\pi(x^2 + y^2)^{\frac{3}{2}}} \right), \quad (7)$$

где  $F_r$  – оператор преобразования Фурье. Моногенный сигнал формально определяется как

$$F_M(u, v) = F(u, v) + i\mathbf{R} \cdot F(u, v), \quad (8)$$

где  $F(u, v)$  – Фурье преобразование  $f(x, y)$  и  $\mathbf{R} = (R_x, R_y)$ .

Для обеспечения инвариантности к масштабированию предлагается использовать линейное представление Пуассона [16], так как оно связано с моногенным сигналом. Это представление  $f_p(x, y, s)$  определяется как свертка изображения  $f(x, y)$  с ядром Пуассона, следующим образом:

$$f_p(x, y, s) = \frac{s}{2\pi(x^2 + y^2 + s^2)} * f(x, y) = F_r^{-1} \left( e^{-2\pi s \sqrt{u^2 + v^2}} \cdot F(u, v) \right), \quad (9)$$

где  $s$  – параметр масштаба, который контролирует разрешение изображения,  $F_r^{-1}$  – оператор обратного преобразования Фурье. Представление Пуассона в частотной области после полосовой фильтрации задается следующим образом:

$$F_{bp}(u, v) = \left( e^{-2\pi s_0 \lambda^k \sqrt{u^2 + v^2}} - e^{-2\pi s_0 \lambda^{k-1} \sqrt{u^2 + v^2}} \right) \cdot F(u, v), \quad (10)$$

где  $\lambda \in (0; 1)$  указывает относительную ширину полосы частот,  $s_0$  – параметр самого грубого масштаба, а  $k \in N$  обозначает номер полосы пропускания [16]. Тогда представление Пуассона моногенного сигнала имеет вид:

$$F_{Mbp}(u, v) = F_{bp}(u, v) + i\mathbf{R} \cdot F_{bp}(u, v), \quad (11)$$

где функции

$$f_p(x, y) = F_r^{-1} \{F_{bp}(u, v)\}, \quad (12)$$

$$f_x(x, y) = F_r^{-1} \{R_x(u, v) F_{bp}(u, v)\}, \quad (13)$$

$$f_y(x, y) = F_r^{-1} \{R_y(u, v) F_{bp}(u, v)\}, \quad (14)$$

заданы в пространственной области. Локальная энергия  $E(x, y)$ , локальная ориентация  $\theta_{or}(x, y)$  и локальная фаза  $\varphi(x, y)$  могут быть вычислены следующим образом:

$$E(x, y) = \sqrt{(f_p(x, y))^2 + (f_x(x, y))^2 + (f_y(x, y))^2}, \quad (15)$$

$$\theta_{or}(x, y) = \arctan\left(\frac{f_y(x, y)}{f_x(x, y)}\right), \quad (16)$$

$$\varphi(x, y) = \arctan\left(\frac{\sqrt{(f_x(x, y))^2 + (f_y(x, y))^2}}{f_p(x, y)}\right). \quad (17)$$

## 2. ПРЕДЛАГАЕМЫЙ ДЕСКРИПТОР ПРИЗНАКОВ

Чтобы обнаружить особые объекты для построения дескрипторов будем использовать модифицированный угловой детектор Харриса [17] для моногенного сигнала:

$$H \cong T_M = \sum_m \sum_n w(m, n) \begin{bmatrix} (f'_x)^2 & f'_x f'_y \\ f'_x f'_y & (f'_y)^2 \end{bmatrix}, \quad (18)$$

где

$$f'_x(x, y) = \frac{f_x(x, y)}{\sum_n A_n(x, y)},$$

$$f'_y(x, y) = \frac{f_y(x, y)}{\sum_n A_n(x, y)},$$

$w(m, n)$  – весовые коэффициенты.

Функция углового детектора используется для получения кандидатов признаков,

$$M_c(x, y) = \det(T_M) - \beta \cdot (\text{trace}(T_M))^2, \quad (19)$$

где  $\beta$  – параметр чувствительности. Обычно используют  $\beta = 0.04$ . Полученные кандидаты используются для выделения особых объектов с высокой степенью конгруэнтности

$$M'_c(x, y) = M_c(x, y) \cdot PC(x, y), \quad (20)$$

Рис. 1 иллюстрирует работу предложенного обнаружителя особых объектов.

Известно, что гистограммы ориентированных градиентов [18, 19] устойчивы к небольшим искажениям сигнала и шуму. Предлагается построить дескриптор на основе гистограммы ориентированных градиентов. После обнаружения особых точек, вокруг каждой точки строится локальная окрестность  $16 \times 16$  элементов, сглаживается гауссовым окном ( $\sigma = 1.5$ ). Далее, окрестность разделяется на области  $4 \times 4$  элементов, а для каждой области вычисляется гистограмма ориентированных градиентов с использованием функции локальной ориентации следующим образом:

$$HOGC(bin_0) = HOGC(bin_0) + PC(x, y), \quad (21)$$

$$bin_0 = \left\lfloor \frac{nbins}{360} \right\rfloor \theta_{or}(x, y),$$

$nbins$  – количество уровней квантования в гистограмме ориентированных градиентов.

Рис. 2 иллюстрирует работу предложенного дескриптора.

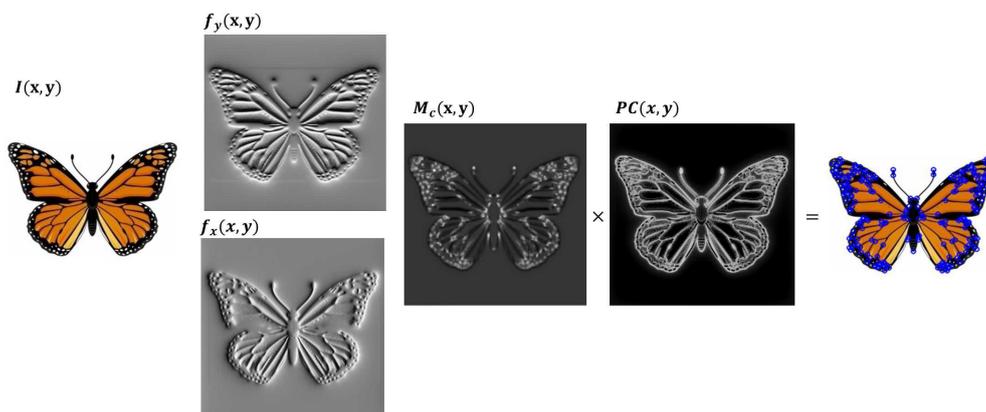


Рис. 1. Обнаружение особых объектов для построения дескриптора признаков.

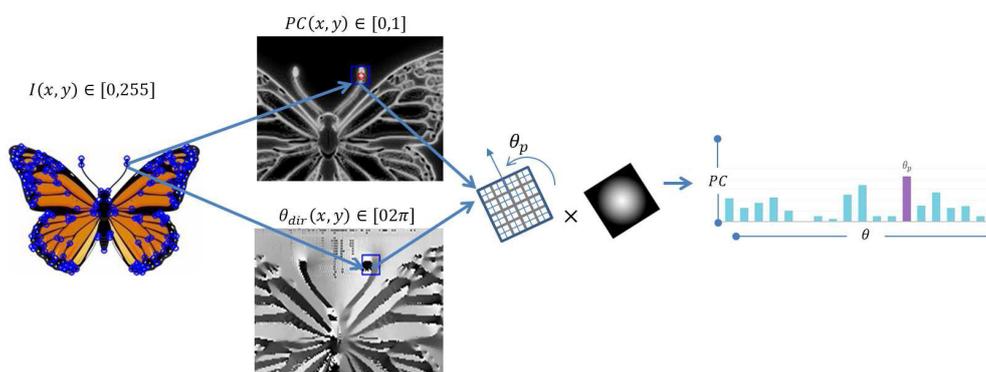


Рис. 2. Построение дескриптора признаков.

### 3. КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ

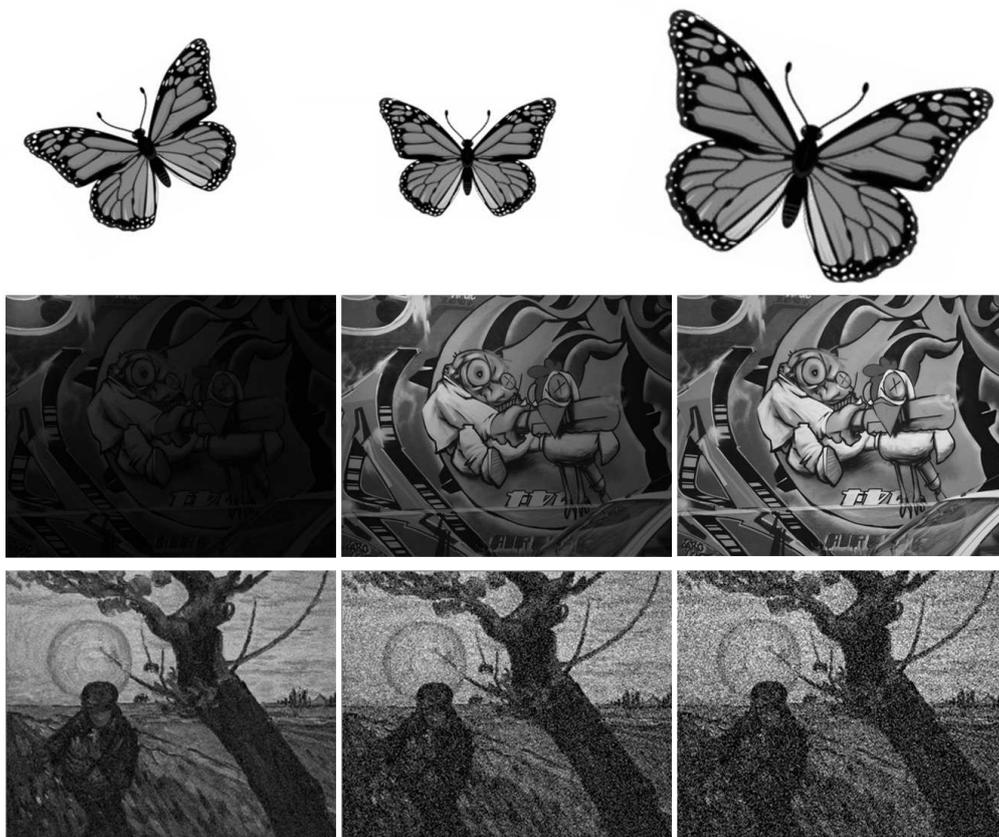
Тестовые изображения содержали символы, искаженные с помощью таких преобразований как поворот в диапазоне  $[-30, 30]$  градусов с шагом 5 градусов, изменение масштаба в диапазоне  $[0.8, 1.3]$  с шагом 0.1, контраста в диапазоне  $C \in [0.5, 2]$  с шагом 0.3. Кроме того, были проведены эксперименты с различными параметрами шума и освещенности. Тестовые сцены были одновременно искажены и аддитивным шумом с шагом среднеквадратичного отклонения  $\sigma$  равным 5. Неравномерное освещение моделировалось с помощью Ламбертовой модели [20]

$$d(\vec{x}) = \cos \left\{ \frac{\pi}{2} - \arctan \left[ \frac{\rho}{\cos(\theta)} \left[ (\rho \tan(\theta) \cos(\varphi) - x_1)^2 + (\rho \tan(\theta) \sin(\varphi) - x_2)^2 \right]^{-1/2} \right] \right\}, \quad (22)$$

где параметры имели следующие значения:  $\theta = 65$ ,  $\varphi = 60$  и  $\rho \in [10, 50]$  с шагом 10.

Критерием оценки качества тестируемых дескрипторов является нормированное количество правильных сопоставлений, которое вычисляется как отношение количество правильных сопоставлений к общему количеству сравнений (matching score (MS)). Предлагаемый алгоритм построения дескрипторов (МНОГ) сравнивался с наиболее известными алгоритмами SIFT [21] и SURF[22]. Общее количество синтезированных изображений более шести тысяч, которые получены из трех разных сцен с помощью геометрических, яркостных изменений, а также зашумления. Примеры тестовых изображений показаны на Рис. 3.

На Рис. 4 представлена сравнительная характеристика работы предлагаемого алгоритма с известными дескрипторными алгоритмами на тестовых искаженных изображениях.



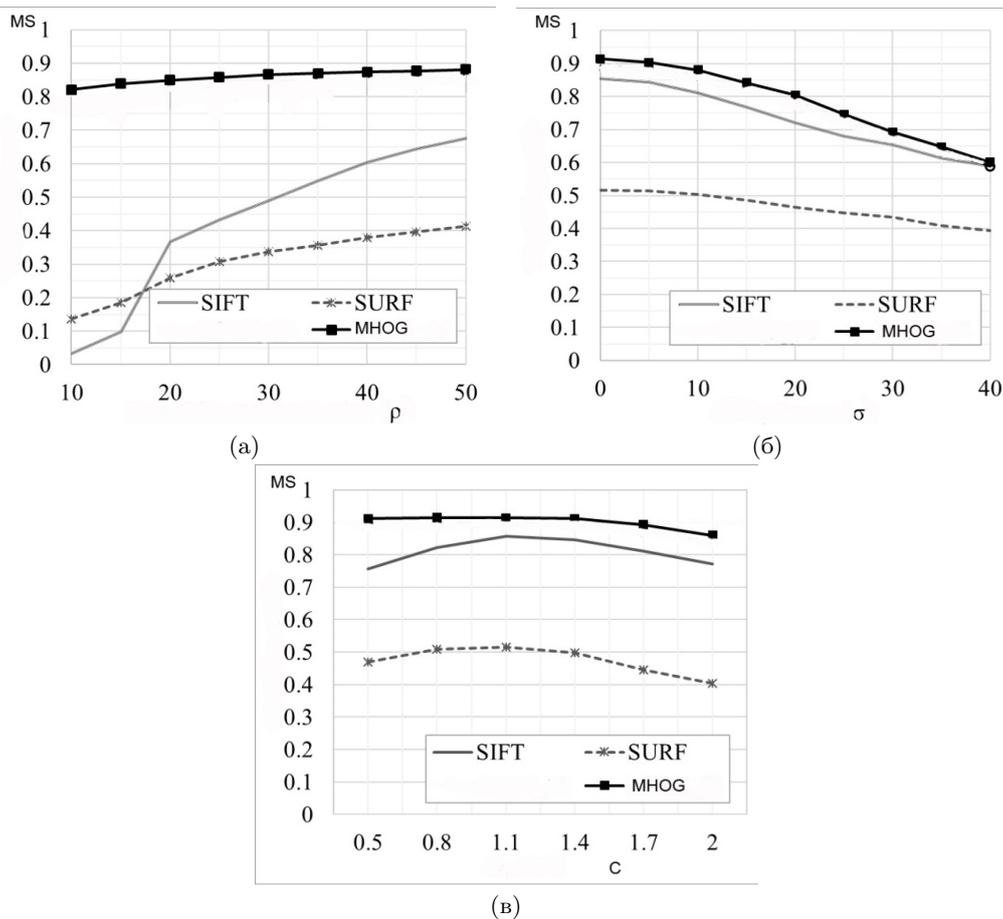
**Рис. 3.** Тестовые изображения: сверху вниз: сцена бабочки после поворота и масштабирования; сцена граффити при различном неравномерном освещении; и тестовая сцена, искаженная аддитивным шумом.

### ЗАКЛЮЧЕНИЕ

В данной работе предложен инвариантный к неравномерному освещению дескриптор для распознавания образов на искаженных изображениях с использованием моногенного сигнала и фазовой конгруэнтности. С помощью компьютерного моделирования была продемонстрирована эффективность предлагаемого дескриптора по сравнению с известными современными дескрипторами. Эксперименты проведены на тестовых сценах, искаженных неравномерной освещенностью, геометрическими преобразованиями и шумом. Поскольку предлагаемый подход к сопоставлению изображений по своей природе является локальным, то его быстрая реализация возможна с использованием графических процессоров и технологий параллельного программирования.

### СПИСОК ЛИТЕРАТУРЫ

1. Alexander Andreopoulos and John K Tsotsos. 50 years of object recognition: Directions forward. *Computer Vision and Image Understanding*, 117(8):827–891, 2013.
2. BVK Vijaya Kumar, Abhijit Mahalanobis, and Richard D Juday. *Correlation pattern recognition*. Cambridge University Press, 2005.
3. Richard O Duda, Peter E Hart, and David G Stork. *Pattern classification*. John Wiley and Sons, 2007.
4. Tinne Tuytelaars, Krystian Mikolajczyk, et al. Local invariant feature detectors: a survey. *Foundations and trends in computer graphics and vision*, 3(3):177–280, 2008.



**Рис. 4.** Сравнение работы дескрипторных алгоритмов на тестовых изображениях с точки зрения нормированного количества правильных сопоставлений MS: а) при неравномерном освещении ( $\rho$ ); б) зашумлении ( $\sigma$ ); в) изменении контраста ( $C$ ).

5. David H Hubel, Janice Wensveen, and Bruce Wick. Eye, brain, and vision. Scientific American Library New York, 1995.
6. Evgeny Gladilin and Roland Eils. On the role of spatial phase and phase correlation in vision, illusion, and cognition. *Frontiers in computational neuroscience*, 9:45, 2015.
7. F. Attneave. Some informational aspects of visual perception. *Psychological review*, 61(3):183, 1954.
8. M Concetta Morrone, John Ross, David C Burr, and Robyn Owens. Mach bands are phase dependent. *Nature*, 324(6094):250–253, 1986.
9. M Concetta Morrone and Robyn A Owens. Feature detection from local energy. *Pattern recognition letters*, 6(5):303–313, 1987.
10. M Concetta Morrone and DC Burr. Feature detection in human vision: A phase-dependent energy model. *Proceedings of the Royal Society of London B: Biological Sciences*, 235(1280):221–245, 1988.
11. Ben Robbins and Robyn Owens. 2d feature detection via local energy. *Image and Vision Computing*, 15(5):353–368, 1997.
12. Peter Kovesi. Image features from phase congruency. *Videre: Journal of computer vision research*, 1(3):1–26, 1999.
13. Peter Kovesi et al. Edges are not just steps. In *Proceedings of the Fifth Asian Conference on Computer Vision*, volume 8, pages 22–8. Melbourne, 2002.

14. Peter Kovési. Phase congruency: A low-level image invariant. *Psychological research*, 64(2):136–148, 2000.
15. Michael Felsberg and Gerald Sommer. The monogenic signal. *IEEE Transactions on Signal Processing*, 49(12):3136–3144, 2001.
16. Michael Felsberg and Gerald Sommer. The monogenic scale-space: A unifying approach to phase-based image processing in scale-space. *Journal of Mathematical Imaging and vision*, 21(1):5–26, 2004.
17. Chris Harris and Mike Stephens. A combined corner and edge detector. In *Alvey vision conference*, volume 15, pages 10–5244. Citeseer, 1988.
18. Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005.
19. Julia Diaz-Escobar and Vitaly Kober. A robust hog-based descriptor for pattern recognition. In *SPIE Optical Engineering+ Applications*, pages 99712A–99712A. International Society for Optics and Photonics, 2016.
20. Victor H Diaz-Ramirez, Kenia Picos, and Vitaly Kober. Target tracking in nonuniform illumination conditions using locally adaptive correlation filters. *Optics Communications*, 323:32–43, 2014.
21. David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.
22. Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. In *European conference on computer vision*, pages 404–417. Springer, 2006.

## **A new feature-invariant feature descriptor for pattern recognition**

**Diaz-Escobar J, Kober V., Karnaukhov V. and Gonzalez-Fraga J.A.**

A new illumination-invariant feature descriptor in color images is proposed. The proposed method is on the local energy model considered as a biologically plausible visual system model. The algorithm of feature detection and descriptor calculation is constructed with the monogenic scale-space and phase congruency approach using modified histograms of oriented gradients. Computer simulation results show that the proposed descriptor yields a superior feature detection and under illumination change, noise, and slight geometric distortions comparing with that of the state-of-the-art descriptors.

**KEYWORDS:** feature descriptor, non-uniform illumination, phase congruency, monogenic signal, histogram of oriented gradients.