

АВТОМАТИЗИРОВАННАЯ СИСТЕМА ШУМОПОДАВЛЕНИЯ С ИСПОЛЬЗОВАНИЕМ МЕТОДА ГЛУБОКОГО ОБУЧЕНИЯ ДЛЯ РАСПОЗНАВАНИЯ ЛИЦ

AUTOMATED IMAGE DENOISING SYSTEM USING DEEP LEARNING METHOD FOR FACE RECOGNITION

Naing Min Tun
A. Gavrilov
Nyan Linn Tun
Pyaе Phyо Paing
Htet Aung

Summary. Analysis of various methods of noise reduction used in the facial recognition system is given. The algorithm of the autoencoder for solving the problem of noise reduction is constructed. Python program is used for modeling the algorithm of the autoencoder. The results of the proposed method have been compared with other standard noise reduction methods. Estimating the efficiency of the algorithm is given using machine learning methods for face recognition purposes.

Keywords: denoising, autoencoder, face recognition, deep learning.

Наинг Мин Тун

Аспирант, Московский государственный
технический университет им. Н.Э. Баумана
naingminhtun52@gmail.com

Гаврилов Александр Игоревич

К.т.н., доцент, Московский государственный
технический университет им. Н.Э. Баумана
alexgavrilov@mail.ru

Ньян Линн Тун

Аспирант, Московский государственный
технический университет им. Н.Э. Баумана
nyanlin54@gmail.com

Пья Пьо Паинг

Аспирант, Московский государственный
технический университет им. Н.Э. Баумана
ppaing12@gmail.com

Хтет Аунг

Аспирант, Московский государственный
технический университет им. Н.Э. Баумана
happyland27057@gmail.com

Аннотация. Приведен анализ различных методов уменьшения помехи, использованный в системе распознавания лиц. Построен алгоритм автоэнкодера решения задачи уменьшения помехи. Использована программа Питон для моделирования рассмотренных подходов. Показаны результаты сравнения предложенного метода с нескольких методов уменьшения помехи. Приведены оценки эффективности работы алгоритма с помощью методов машинного обучения для целей распознавания лиц.

Ключевые слова: шумоподавление, авторэнкодер, распознавание лиц, глубокое обучение.

Введение

В настоящее время распознавание лиц достигло большого прогресса для различных потенциальных применений в областях безопасности и чрезвычайных ситуаций [1–4], правоохранительных органов [5] и видеонаблюдения [6–8], контроля доступа [9] и т.д. Однако в некоторых неконтролируемых условиях, включая различное освещение, позы, мимику и шум, производительность системы распознавания лиц будет резко снижена. Были проведены обширные работы по решению проблем освещения, ракурса, выражения лица и получены положительные результаты [10–12]. Однако изображение лица уязвимо для шумов при его получе-

нии, квантовании, сжатии и т.д.; когда рассматриваются зашумленные изображения, точность распознавания большинства подходов значительно падает.

Изображение часто искажается шумом при его получении и передаче. Шумоподавление используется для удаления аддитивного шума, сохраняя при этом как можно больше важных характеристик сигнала. Как правило, наборы данных, собираемые датчиками изображения, загрязнены шумом. Несовершенные инструменты, проблемы с процессом сбора данных и вмешивающиеся природные явления-все это может повредить интересующим данным. Таким образом, шумоподавление является важной технологией в анализе изображений

и первым шагом, который необходимо сделать перед анализом изображений. Поэтому для предотвращения такого рода искажений цифровых изображений необходимы методы шумоподавления [13].

Шум содержит нежелательную информацию в цифровых изображениях и создает нежелательные эффекты, такие как артефакты, нереалистичные края, невидимые линии, углы, размытые объекты, нарушает фоновые сцены. Чтобы уменьшить эти нежелательные эффекты, предварительное изучение моделей шума необходимо для дальнейшей обработки. Существует несколько моделей шума, некоторые из которых представлены в работе [14].

Были предложены различные методы снижения влияния шума на изображения до стадии распознавания. Одним из подходов является трансформация сигналов изображения в альтернативную область, где они могут быть более легко отделены от шума [15–17]. Другая группа методов — это захват статистики изображений непосредственно в области изображений [18, 19]. Обе эти категории подходов позволяют получить изображения хорошего качества. Но изображение после подавления помех имеет тенденцию терять часть информации, что вредит распознаванию изображения на последующей стадии. С развитием глубоких искусственных нейронных сетей может быть достигнут сквозной процесс шумоподавления. В этой статье мы используем автоэнкодер для достижения шумоподавления.

Методы шумоподавления

Существующие методы способны дать хорошие результаты во многих практических сценариях. Среди стандартных методов шумоподавления можно выделить медианную фильтрацию, байесшринк, висушринк и автоэнкодер.

Медианный фильтр: медианный фильтр работает по принципу скользящего окна. Ядро размером 3×3 , 5×5 или 7×7 пикселей сканируется поверх пиксельной матрицы всего изображения. Вычисляется медиана значений пикселей в окне, и центральный пиксель окна заменяется вычисленной медианой. Медиана более устойчива по сравнению со средним значением. Таким образом, одиночный нерепрезентативный пиксель в окрестности не будет существенно влиять на медианное значение. Поскольку медианное значение должно фактически быть значением одного из пикселей в окрестности, медианный фильтр не создает новых нереалистичных значений пикселей, когда фильтр пересекает границу области. По этой причине медианный фильтр гораздо лучше сохраняет острые края, чем усредняющий фильтр.

Байесшринк: Байесшринк был предложен Чангом, Ю и Веттерли [20]. Цель этого метода заключается в минимизации байесовского риска, поэтому метод получил название Байесшринк. Он использует мягкое пороговое значение и зависит от поддиапазона, что означает, что пороговое значение выполняется в каждой полосе разрешения в вейвлет-декомпозиции. Порог Байеса T_B определяется как:

$$T_B = \frac{\sigma^2}{\sigma_X}$$

где σ^2 — дисперсия шума, а σ_X — дисперсия сигнала без шума. Выход метода Байесшринка гораздо ближе к высокому качеству изображения.

Висушринк: Висушринк был введен Донохо [21]. Он использует пороговое значение T , пропорциональное стандартному отклонению шума. Он следует жесткому правилу порогового удержания и определяется как:

$$T = \sigma \sqrt{2 \log(n)}$$

где σ — стандартное отклонение шума от уровня шума, а n — размер выборки.

Висушринк можно рассматривать как применение универсальных пороговых селекторов, которые проявляют близкие к оптимальным свойства минимаксной ошибки и обеспечивают с высокой вероятностью, что оценки так же гладки, как истинные базовые функции [21]. Однако Висушринк, как известно, дает восстановленные изображения, которые чрезмерно сглажены. Это происходит потому, что Висушринк удаляет слишком много коэффициентов. Другим недостатком является то, что он не может удалить спекл-шум (speckle noise). Он может иметь дело только с аддитивным шумом. Висушринк следует схеме глобального порогового значения [22], где существует одно пороговое значение, применяемое глобально ко всем вейвлетам.

Автоэнкодер — это тип искусственной нейронной сети, используемый с целью сжатия данных или удаления шума. Некоторые исследователи отметили, что автоэнкодер имеет возможность изучать основные представления цифровых сигналов, и дает возможность работы с сетевой структурой, только если ограничить ожидаемый выходной сигнал таким же или аналогичным входному. Кроме того, автоэнкодер может эффективно восстанавливать изображения [23].

Авторэнкодер имеет основные свойства, которые зависят от данных, с потерями и могут автоматически изучать представления ядра из входных примеров без каких-либо контрольных сигналов, следовательно, метод относится к обучению без учителя [24]. Более того, в ав-



Рис. 1. Функционирование автоэнкодера по шумоподавлению изображения.

торэнкодере энкодер и декодер обычно являются нейронными сетями. Автоэнкодер зависит от данных, так как сеть может использоваться только для сжатия данных, аналогично тому, как они были обучены, поэтому эта структура зависит от конкретной задачи. Автоэнкодер также является методом сжатия с потерями, т.е. выход может иногда иметь плохую производительность. Автоэнкодер настраивается автоматически на примерах данных и представляет собой сквозной процесс обучения.

В этой статье автоэнкодер использован для уменьшения шума изображения. В общем случае он состоит из двух основных частей: кодера и декодера. Кодер отображает входные изображения в скрытое представление и декодер восстанавливает оригинал изображения из кодов, генерируемых кодером. Типичный рабочий процесс сети сверточного автоэнкодера показан на рис. 1.

Как математическое выражение, автоэнкодер берет незамеченный обучающий входной набор данных (домен X) и создает изображения (домен F). Общее описание автоэнкодера можно представить следующим образом:

$$\begin{aligned} \phi : X &\rightarrow F \\ \psi : F &\rightarrow X \\ \phi, \psi &= \arg \min_{\phi, \psi} \|X - (\psi \circ \phi) X\|^2 \end{aligned}$$

Где, ϕ обозначает функцию кодера, которая отображает входные данные изображения X в скрытое пространство (скрытые представления) F . ψ обозначает функцию декодера, которая отображает скрытое пространство F на выход X .

Сеть кодировщика автоэнкодера может быть представлена стандартной функцией нейронной сети через функцию активации следующим образом:

$$z = \sigma(Wx + b)$$

где z — элемент скрытого пространства. Аналогично, сеть декодера может быть представлена в той же форме, но с другим весом, смещением и используемой функцией активации.

$$x' = \sigma'(W'z + b')$$

Где, σ', W', b' сети декодера могут быть не связаны σ, W, b сети энкодера.

Автоэнкодеры обучены минимизировать функцию потерь, которая может быть записана в терминах функции нейронной сети с использованием стандартного метода обратного распространения. Это можно записать следующим образом:

$$L(x, x') = \|x - x'\|^2 = \|x - \sigma'(W'(\sigma(Wx + b)) + b')\|^2$$

Где, x обычно усредняется по некоторому входному обучающему набору.

Решение задачи шумоподавления

Необходимо выполнить этап шумоподавления для того, чтобы сделать систему распознавания по зашумленным изображениям такой, как описано выше. Также, необходимо применить модель шума к изображениям для оценки моделей шумоподавления. Упрощенная общая модель деградации изображения для задачи шумоподавления, широко принятая в литературе, определяется как:

$$y = x + n$$

где x относится к неизвестному высококачественному изображению (основная истина), y — ухудшенное изображение, а n представляет аддитивный шум.

В данной работе рассматривается гауссов шум в виде следующего уравнения:

$$P(g) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(g-\mu)^2}{2\sigma^2}}$$

где g — серый уровень, σ — стандартное отклонение и μ — среднее значение. Результаты добавления гауссова шума представлены на рис. 2.

После применения гауссовой модели шума к изображениям, создается модель автоэнкодера с целью

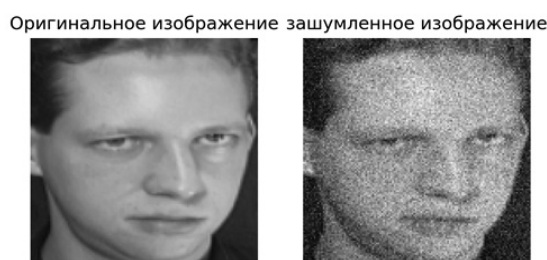


Рис. 2. Изображение до и после добавления шума

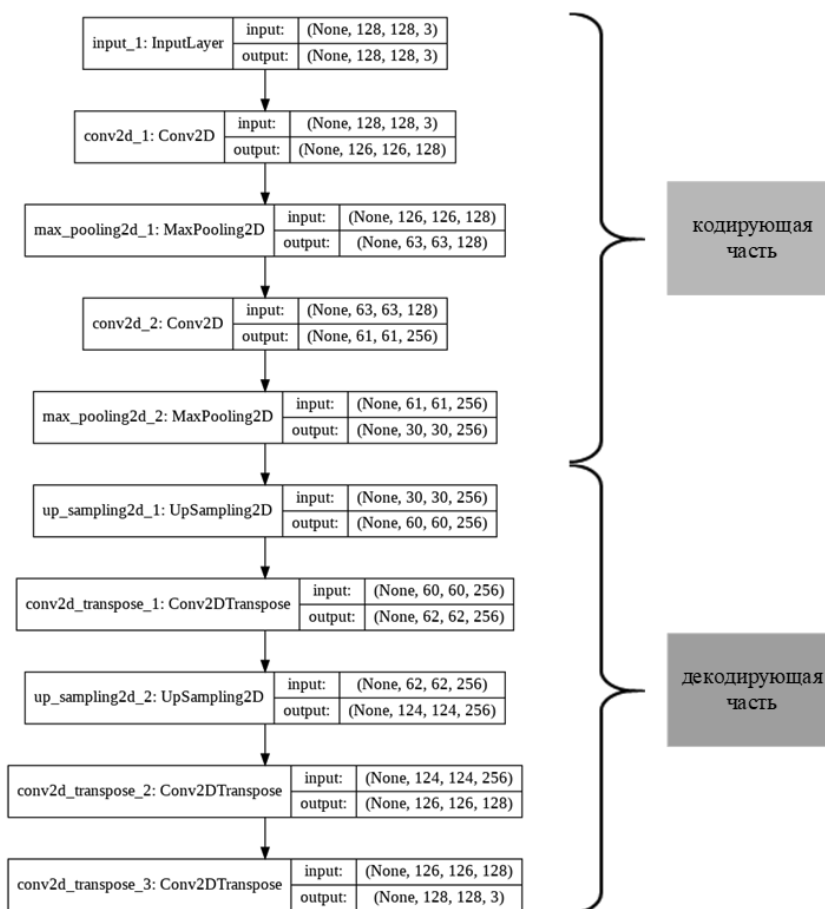


Рис. 3. Сетевая архитектура свёрточного авторэнкодера

устранения шума из изображений. Архитектура автоэнкодера составлена из двух 2D сверточных слоев для кодирующей части и двух 2D деконволюционных слоя для декодирующей части. Архитектура модели автоэнкодера представлена на рисунке 3. Модель автоэнкодера обучена в системе облачной службы «Google Colab», поддерживаемой GPU. ORL лицевая база данных, используемая для эксперимента, состоит из 400 изображений размером 112x92 и 256 уровней серого цвета на пиксель. Эта база данных разделена на две группы с целью обучения модели автоэнкодера, 75% для обуча-

ющей части и 25% для тестовой части соответственно. Потери на одну эпоху в процессе обучения представлены на рисунке 4.

Результаты шумоподавления на случайно выбранном тестовом изображении из ORL базы данных представлены на рис. 5 и в таблице 1.

Также ясно, что производительность автоэнкодера лучше, чем у других методов при различных уровнях шума.

Таблица 1. Сравнение результатов PSNR с использованием различных методов шумоподавления

значение параметра помехи	С шумом	Медианный метод	Байесшринк	Висушринк	Авторэнкодер
$\sigma = 0.10$	19.99	29.80	29.73	26.69	31.91
$\sigma = 0.15$	16.45	26.62	27.89	23.73	29.88
$\sigma = 0.20$	14.03	24.50	26.84	21.67	27.35

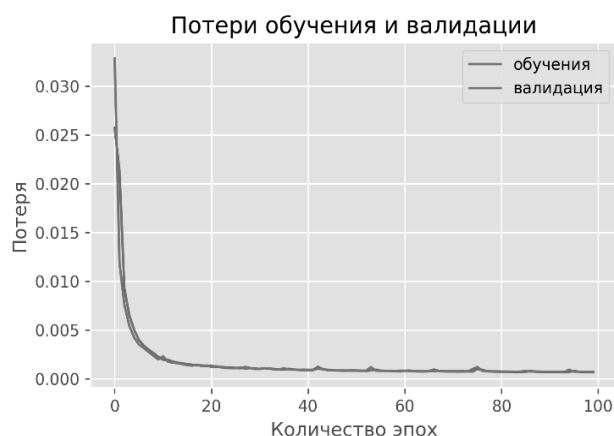


Рис. 4. Обучение авторэнкодера

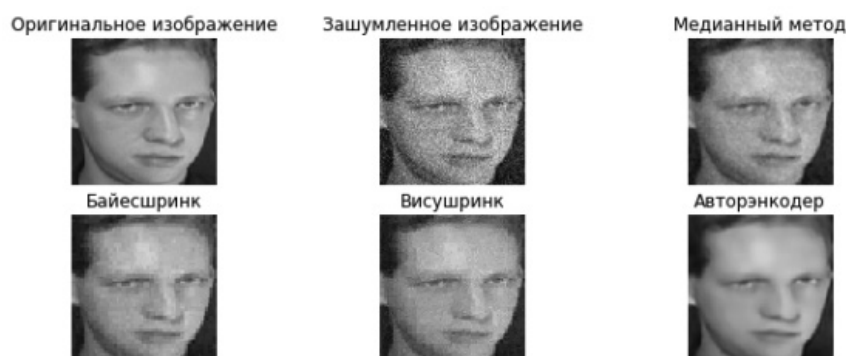


Рис. 5. Сравнение результатов на зашумленном изображении с использованием различных методов шумоподавления

Для распознавания лиц использованы модель FaceNet [25] и метод опорных векторов (SVM). Проведено исследование как эта модель распознавания лиц может работать на зашумленных изображениях с различными фильтрами шумоподавления или без них. Результаты представлены в виде кривой precision-recall на рис. 6, а оценки точности при различной плотности шума $\sigma = 0.10, 0.15$ и 0.20 приведены в таблице 2.

В соответствии с рисунком 6 и таблицей 2 можно сделать вывод, что предлагаемая сверточная модель автоэнкодера может обеспечить лучшую производи-

тельность модели распознавания лиц, чем другие фильтры шумоподавления.

Заключение

В работе представлен метод шумоподавления на изображениях на основе сверточного автоэнкодера. Предложены архитектура автоэнкодера, принимающей на вход зашумленные двумерные изображения разрешением 128×128 пикселей и отдающей изображения без шумов с тем же самым разрешением. Предложенная архитектура была реализована с использованием

Таблица 2. Результаты точности описаны по различной плотности шума

Плотность шума	Без шумов	С шумом	Медианный метод	Байесшринк	Висушрик	Авторэнкодер
$\sigma = 0.10$	99.46	97.56	99.35	99.04	98.90	99.38
$\sigma = 0.15$	99.46	88.43	97.72	95.28	94.79	99.33
$\sigma = 0.20$	99.46	63.25	96.09	84.35	85.26	97.49

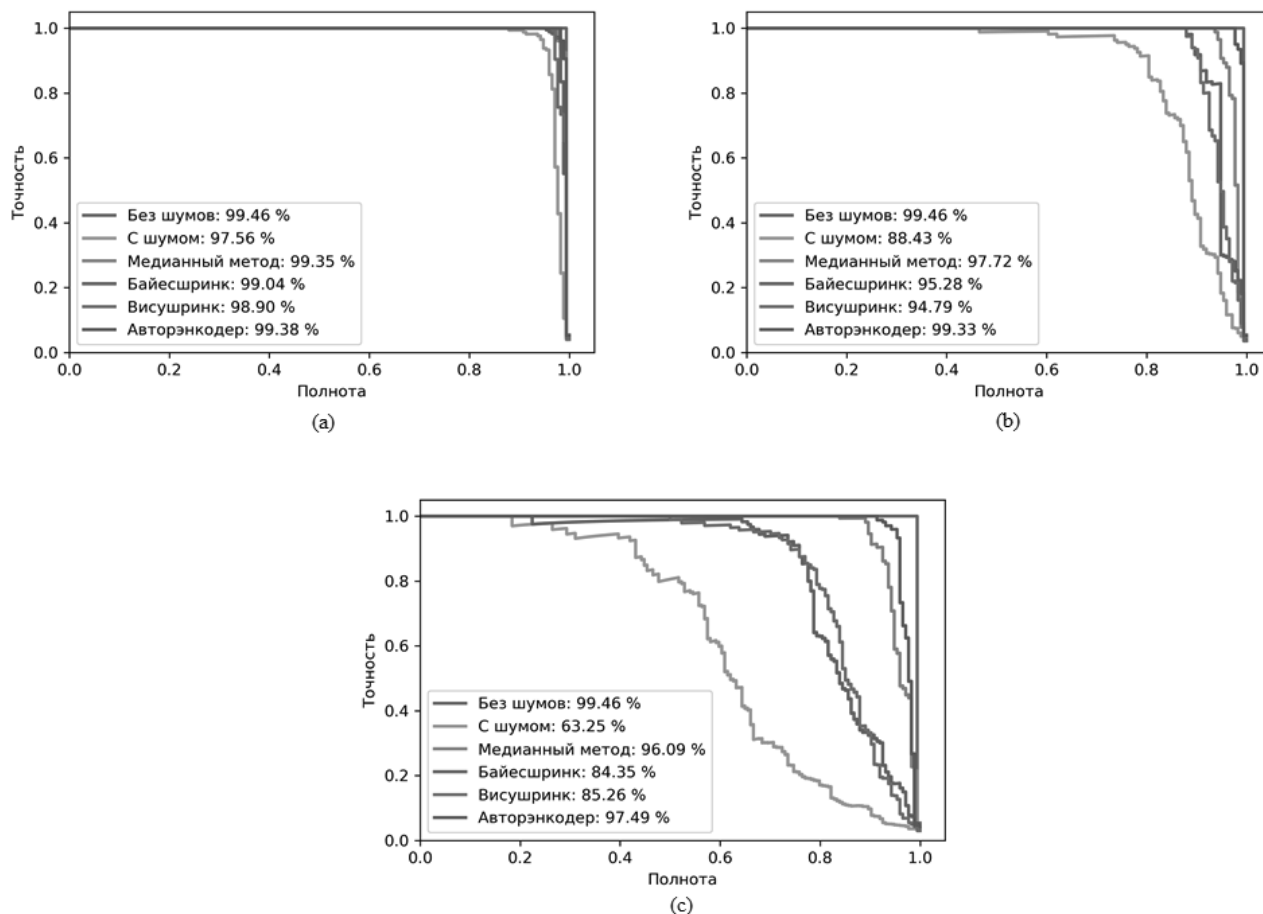


Рис. 6. Оценки точности распознавания лиц при различных плотностях шума (а) плотность гауссовского шума $\sigma = 0.10$ (б) $\sigma = 0.15$ (в) $\sigma = 0.20$

библиотеки Keras. Обучение произведено с использованием графического процессора GPU на Google Colab. Тестирование архитектуры показало, что она хорошо обучается на зашумленных изображениях. Произведено сравнение предложенного метода автоэнкодера с несколькими стандартными методами шумоподавления и утверждено то, что наилучшим методом является авторэнкодер по факту пикового отношения сигнал / шум.

Основным преимуществом алгоритма автоэнкодера является возможность качественного восстановления из зашумленных изображений. Основным недостатком

представленного алгоритма автоэнкодера является необходимо использовать графические процессоры видеокарт по причине высокой вычислительной сложности.

В работе рассмотрено применение автоэнкодера в задаче распознавания лиц путем предварительной обработки фильтрации. Представлено влияние автоэнкодера на систему распознавания лиц при условии зашумленных обстоятельствах. Одна из возможных будущих работ заключается в использовании автоэнкодера в качестве этапа предварительной обработки системы распознавания лиц в ночное время для дальнейшего повышения точности и надежности.

ЛИТЕРАТУРА

1. Padmapriya S., Kalajames E. A., Real Time Smart Car Lock Security System Using Face Detection and Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012, pp. 1–6
2. Z. Xu, C. Hu, L. Mei, Video structured description technology-based intelligence analysis of surveillance videos for public security applications. *Multimed. Tools Appl.* 75(19), 1–18 (2015)
3. Z. Xu, Y. Liu, H. Zhang et al., Building the multi-modal storytelling of urban emergency events based on crowdsensing of social media analytics. *Mob. Netw. Appl.* 22(2), 218–227 (2017)
4. Y. Yang, Z. Xu et al., A security carving approach for AVI video based on frame size and index. *Multimedia Tools Appl.* 76(3), 3293–3312 (2017)
5. D. McAllister, Law Enforcement Turns to Face-Recognition Technology. *Information Today.* 24(5) (2007)
6. Z. Yan, Z. Xu, JD., The Big Data Analysis on the Camera-based Face Image in Surveillance Cameras. *Intell. Autom. Soft Comput.* doi: 10.1080/10798587.2016.1267251 (2016)
7. Z. Xu, et al., The big data analytics and applications of the surveillance system using video structured description technology. *Clust. Comput.* 19(3), 1283–1292 (2016)
8. B. Kamgarpari, W. Lawson, B. Kamgarpari, Toward development of a face recognition system for watchlist surveillance. *IEEE Trans. Pattern Anal. Mach. Intell.* 33(10), 1925–1937 (2011)
9. S.J. McKenna, S. Gong, Non-intrusive person authentication for access control by visual tracking and face recognition. *Lect. Notes Comput. Sci.* 1206, 177–183 (2006)
10. H. Roy, D. Bhattacharjee, Local-gravity-face (LG-face) for illumination-invariant and heterogeneous face recognition. *Info. Forensics Secur. IEEE Trans.* 11(7), 1–1 (2016)
11. X Wang, Q Ruan, Jin, et al., Three-dimensional face recognition under expression variation. *EURASIP J. Image Video Process.* 54(1): 1–11 (2014)
12. MH Siddiqi et al., Human facial expression recognition using curvelet feature extraction and normalized mutual information feature selection. *Multimedia Tools Appl.* 75(2), 935–959 (2016)
13. Rajni, Anutam, "Image Denoising Techniques –An Overview," *International Journal of Computer*, Vol. 86, No.16, January 2014.
14. Ajay Kumar Boyat, Brijendra Kumar Joshi, "A REVIEW PAPER: NOISE MODELS IN DIGITAL IMAGE PROCESSING," *Signal & Image Processing: An International Journal (SIPIJ)* Vol.6, No.2, pp.63–75, 2015
15. J Xu, K Zhang, M Xu, Z Zhou, "An adaptive threshold method for image denoising based on wavelet domain," *Proc. SPIE Int. Soc. Opt. Eng.* 7495, 165 (2009)
16. J Portilla, V. Strela et al., Image denoising using scale mixtures of Gaussians in the wavelet domain. *IEEE Trans. Image Process.* 12(11), 1338–1351 (2003)
17. F Luisier, T. Blu, M Unser, A new SURE approach to image denoising: interscale orthonormal wavelet thresholding. *IEEE Trans. Image Process.* 16(3), 593–606 (2007)
18. B.A. Olshausen, D. J. Field, Sparse coding with an overcomplete basis set: a strategy employed by V1? *Vision Res.* 37(23), 3311–3325 (1997)
19. M Elad, M Aharon, Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. Image Process.* 15(12), 3736–3745 (2006)
20. S. Grace Chang, Bin Yu and Martin Vetterli, "Adaptive Wavelet Thresholding for Image Denoising and Compression," *IEEE Trans. Image Processing*, Vol 9, No. 9, Sept 2000, pg 1532–1546.
21. David L. Donoho, "De-noising by soft-thresholding," <http://citeseer.nj.nec.com/cache/papers/cs/2831http://zSzzSzwwwwstat.stanford.eduzSzreportszSzdonohezSzenoiserelease3.pdf/donoho94denoising.pdf>, Dept of Statistics, Stanford University, 1992.
22. Anestis Antoniadis, Jeremie Bigot, "Wavelet Estimators in Nonparametric Regression: A Comparative Simulation Study," *Journal of Statistical Software*, Vol 6, I 06, 2001.
23. «Image recovery Theory and application», *Automatica*, vol. 24, no. 5, pp. 726–727, 1988.
24. M. Celebi and K. Aydin, Unsupervised learning algorithms.
25. Florian Schroff, Dmitry Kalenichenko, and James Philbin, "Facenet: A unified embedding for face recognition and clustering," in *CVPR*, 2015, pp. 815–823 *Transl. J. Magn. Japan*, vol. 2, pp. 740–741