UNIVERSITÀ DEGLI STUDI DI PALERMO
Dottorato in Ingegneria dell’Innovazione Tecnologica
Dipartimento di Ingegneria
SSD: ING-INF/05

METHODS AND TECHNIQUES FOR MULTI-SOURCE DATA ANALYSIS AND FUSION

IL DOTTORE
FRANCESCO GUGLIUZZA

IL COORDINATORE
CH.MO PROF. ANTONIO CHELLA

IL TUTOR
PROF. ROBERTO PIRRONE

CICLO XXXI
ANNO CONSEGUIMENTO TITOLO 2019

I dedicate this thesis to my parents, Antonino Gugliuzza and Santa Cammarata, and my closest friends, the best ones a person could have: Gaspare Marotta, Leonardo Menegon and Luca Torrisi.

A special thought goes to my colleague and friend, Dr. Ing. Alessandro Bruno, who continuously inspired me with his character and competence, encouraged me to do better and reassured me with kindness and humor.

Abstract

This work has been inspired by the recent trend in remote sensing and environmental data acquisition. Remote sensing techniques allow us to measure information about an object without touching it. In the last decades remote sensing via satellites has been used in various applications such as Earth observation, weather and storm predictive analysis, atmospheric monitoring, climate change, human-environment interactions. Sensors on airborne and satellite platforms have been recording signals from space for many years, giving rise to a huge amount of data. Some data are processed on-board but others are treated and post-processed in ground stations. Signal and image processing are widely applied on data coming from satellites to extract meaningful information for the aforementioned tasks. Satellites and ground stations communicate with each others by using several transceivers and techniques.

After acquisition, data has to be processed and correlated to generate forecasts and detect unusual phenomena. Human attention works very well when detecting salient patches in an image, and this skill is frequently used to segment areas of interest in satellite images. Saliency is the quality that makes certain regions of an image stand out from the rest of the visual field and grab our attention. Due to the relevance of visual saliency in remote imaging and its applicability to many other scientific and psychological research areas, the main part of this work focuses on the study of theoretical saliency models, the research of novel methods for saliency detection in images and the characterization of human attention on sub-object details. As an extension of these, investigated the effects of Colour Vision Deficiencies (CVDs) on attention and new techniques to alter image colours depending on saliency values have been investigated, in order to restore fixation point coherence between CVD affected and normal observers.

In this work we propose two remote sensing systems, composed by low cost hardware components and portable software solutions, designed to receive meteorological satellite signals. We aim at sampling and processing of the modulated signal entirely in software enabled by Software Defined Radios (SDR) and CPU computational speed overcoming hardware limitations such as high receiver noise and low ADC resolution. For this purpose we developed two software-hardware integrated systems able to perform the following steps: satellite pass prediction, time scheduling, signal demodulation, image cropping and filtering. Then, cloud detection is performed by using two well known clustering algorithms, Otsu and K-means. Although we employed low cost components, we obtained good results in terms of signal demodulation, synchronization and image reconstruction.

In order to emulate human attention to detect and segment salient patches in images (such as satellite images) we have developed two saliency algorithms based on Keypoint Density Maps. Our algorithms work by analyzing the spatial density of keypoints detected in images converted to perceptually uniform color spaces (CIE L*a*b* and CIE L*u*v*); both scale-aware and multi-scale approaches have been implemented in our methods. Furthermore, we compared our methods against the most important and popular saliency detection methods in the state of the art. It is shown that our approaches achieve good results in all experiments. As a reference for performance evaluation, we collected a dataset of eye fixations on objects and sub-object details using eye-tracking technology; the dataset is called Eye-Tracking Through Objects (ETTO) and it is publicly available.

As already said, we developed a technique to use the above-mentioned saliency algorithms to re-color images for individuals affected by color vision deficiencies. Color vision deficiencies affect visual perception of colors and, more generally, color images. We eye-tracked the human fixations in first three seconds of observation of color images to build real fixation point maps, then we detected the main differences between the aforementioned human visual systems related to color vision deficiencies by analyzing real fixation maps among people with and without color vision deficiencies. In this work we provide a method to enhance color regions of the image by using a detailed color mapping of the segmented salient regions of the given image. The segmentation is performed by using the saliency difference between the original input image and the corresponding color blind altered image. A second eye-tracking of color blind people with the images enhanced by using re-coloring of segmented salient regions reveals that the real fixation points are then more coherent (up to 10%) with the normal visual system. The eye-tracking data collected during our experiments are in a publicly available dataset called Eye-Tracking of Colour Vision Deficiencies (EToCVD).

Acknowledgments

I want to thank Basilio Fabio Ceraolo and Calogero Carlo Giunta for the assistance given on the acquisition and processing respectively of the ETTO and the EToCVD datasets.

Contents

Abstract ii

Acknowledgments v

Contents vi

List of Figures viii

List of Tables xii

1 Introduction 1
- 1.1 Goals and Scope 1
- 1.2 Satellite Remote Sensing 2
- 1.3 Visual Saliency 3
- 1.4 Contributions 5
- 1.5 Dissertation Outline 6
- 1.6 Publications 7

2 Satellite Remote Sensing 9
- 2.1 State of the Art 9
  - 2.1.1 Satellite Communications and Signal Reception 9
  - 2.1.2 Cloud Detection 14
- 2.2 Low Resolution Method 16
  - 2.2.1 APT Signal 16
  - 2.2.2 Hardware 17
  - 2.2.3 Software 20

2.2.4 Experimental Results 27
2.3 High Resolution Method 30
2.3.1 HRPT Signal 31
2.3.2 Hardware 31
2.3.3 Software 35

3 Visual Saliency **38**
3.1 State of the Art 38
3.1.1 Eye-tracking Technologies and Datasets 44
3.1.2 Image Content Enhancement through Salient Regions Segmentation for People with Color Vision Deficiencies 46
3.2 Eye-Tracking Through Objects (ETTO) Dataset 48
3.3 Color SIFT Density Map Method 51
3.3.1 HSV Color Space Saliency 52
3.3.2 CIE L*a*b* Color Space Saliency 53
3.3.3 Experimental Results 55
3.4 Multi-scale Color SURF Keypoint Density Map Method 57
3.4.1 Keypoint Density Map 57
3.4.2 Multi-scale KDMs and KDM Improvements 58
3.4.3 Experimental Results 62
3.5 Image Content Enhancement through Salient Regions Segmentation for People with Color Vision Deficiencies 73
3.5.1 Eye-tracking Session 73
3.5.2 Proposed Method 78
3.5.3 Experimental Results 80

4 Conclusions **92**
4.1 Satellite Remote Sensing 92
4.2 Visual Saliency 93
4.2.1 Image Content Enhancement through Salient Regions Segmentation for People with Color Vision Deficiencies 94

Bibliography **96**

List of Figures

2.1 (a) APT frame and (b) row format. 10

2.2 Examples of (a) a multi-channel SAR image acquired by a Space Shuttle and enhanced by polarimetry (NASA/Jet Propulsion Laboratory, 1999), (b) a two-channel NOAA APT image (Group for Earth Observation, 2003) and (c) one channel of a NOAA HRPT image (Medri, 2005). We are interested in images (b) and (c): in (b) each pixel covers an area of 4 km x 4 km, while in (c) each pixel covers an area of 1.1 km x 1.1 km. 13

2.3 (a) The turnstile antenna built for 137 MHz reception, (b) the modified TV amplifier, (c) a RTL2832U-based DVB-T dongle with an Elonics E4000 tuner. 18

2.4 Digital down-converter. 21

2.5 (a) FM demodulator, (b) PLL-based AM demodulator, (c) Envelope detector AM demodulator. 23

2.6 Synchronization block. 25

2.7 Pixel dynamic range calibration block. 26

2.8 AVHRR image after automatic cropping. 27

2.9 Disk filtered remote sensing image from NOAA. 28

2.10 (a) Cloud segmentation obtained with Otsu method and (b) cloud segmentation achieved using the K-means algorithm ($K = 3$). . . . 29

2.11 (a) The offset parabolic antenna, (b) theoretical gain vs. diameter of a parabolic antenna having an efficiency of 0.75. 32

2.12 The helical feed. 33

2.13 (a) The amplifier's schematic, (b) its gain vs. frequency plot, (c) the populated PCB. 34
2.14 Example flowgraph for HRPT reception in GNU Radio (The GNU Radio Foundation, 2009). 36

3.1 The setup used for eye fixation data acquisition. 50
3.2 (a) An image from the OPED and its fixation map (b) before and (c) after Gaussian blurring. 51
3.3 (a) An image from the OPED and (b) its SIFT saliency map calculated in HSV space. 54
3.4 (a) An image from the OPED and (b) its SIFT saliency map calculated in L*a*b* space. 54
3.5 Performance graphs of various saliency models in terms of (a) F-measure vs. threshold and (b) NSS values. 56
3.6 The block diagram of the overall steps of saliency map extraction for a given RGB input image. 61
3.7 (a) An image from the OPED and (b) its saliency map calculated in L*u*v* space using the SURF keypoint detector, multi-scale map generation, polynomial fusion and center bias weighting. 62
3.8 Performance graphs of some variations of our saliency model on the sample dataset. 65
3.9 (a) An image from the ETTO dataset, (b) its fixation map, (c) its saliency map generated with our multi-scale color SURF KDM method, and (d) its saliency map generated with the eDN method. 68
3.10 (a) An image from the MIT1003 dataset, (b) its fixation map, (c) its saliency map generated with our multi-scale color SURF KDM method, and (d) its saliency map generated with the SAM-ResNet method. 69
3.11 (a) An image from the CAT2000 dataset, (b) its fixation map, (c) its saliency map generated with our multi-scale color SURF KDM method, and (d) its saliency map generated with the SAM-VGG method. 70

3.12 (a) An image from the EyeCrowd dataset, (b) its fixation map, (c) its saliency map generated with our multi-scale color SURF KDM method, and (d) its saliency map generated with the Itti-Koch-Niebur method. 71
3.13 (a) An image from the FIGRIM dataset, (b) its fixation map, (c) its saliency map generated with our multi-scale color SURF KDM method, and (d) its saliency map generated with the GBVS method. 72
3.14 Unlike people with normal vision system, people with dichromatic vision system are able to easily recognize the word "NO". 73
3.15 Images taken during the eye-tracking session: starting from the calibration (the far-left image) the eye-tracker records the eye movements, the saccadic movements and the scanpaths. 74
3.16 RGB to CIE L*a*b* conversion allows us to manage with color mapping within color frequencies well perceived by color blind people. 75
3.17 The visual perception of an image can be represented by the fixation points (red diamonds overlaid on the images) for both normal vision system (left column) and color blind vision system (right column). Some details are missed by people with color vision deficiencies and this is revealed by the lack of fixation points on the details noticed by people with normal vision system. 77
3.18 (b) The saliency error is computed as the difference of the saliency maps of (a) the original image and the color blind version of the image. (c) The saliency error regions are segmented and color boosted on CIE L*a*b* color space by using the opposite value with the a* and b* channels, (d) the enhancement is also mapped in the color blind domain. 79
3.19 Perceptual differences highlighting. For a given image (left) some enhancement methods use the average of the differences of L*a*b* channels between the original image and the version with simulated color vision deficiency (center). We adopted the difference of L*a*b* channels between the original image and the version with simulated color vision deficiency, weighted by the saliency difference (right). . 80

3.20 The fixation points (red diamonds overlaid on the images) of observers with normal color vision system (left column) and the fixation points of people affected by protanopia. The images from the right column are from the second eye-tracking session when observing the images enhanced by our method. 82
3.21 For (a) a given image we collected the fixation points from (b) a normal observer, (c) an observer affected by protanopia and (d) an observer affected by deuteranopia. 83
3.22 The enhancement assessment on (a, b) the images is supported by the fixation point maps for observers with (c) protanopia and (d) deuteranopia. 84
3.23 Average AUC score increase of the best 10, 20, 30, 50 cases. Protanopia case study. 86
3.24 Average NSS score increase of the best 10, 20, 30, 50 cases. Protanopia case study. 87
3.25 Average AUC score increase of the best 10, 20, 30, 50 cases. Deuteranopia case study. 88
3.26 Average NSS score increase of the best 10, 20, 30, 50 cases. Deuteranopia case study. 89
3.27 For a given image (first column) we collected the fixation points from normal observer (second column), from observers affected by protanopia (third column), from observers affected by protanopia looking at the enhanced image (fourth column). 90
3.28 For a given image (first column) we collected the fixation points from normal observers (second column), from observers affected by deuteranopia (third column), from observers affected by deuteranopia looking at the enhanced image (fourth column). 91

List of Tables

2.1 Objective metrics results for image enhancement filtering. 30

3.1 Results in various metrics on ETTO dataset. 65
3.2 Results in various metrics on MIT1003 dataset. 66
3.3 Results in various metrics on CAT2000 dataset. 66
3.4 Results in various metrics on EyeCrowd dataset. 67
3.5 Results in various metrics on FIGRIM dataset. 67

Chapter 1

Introduction

1.1 Goals and Scope

My Ph.D. research has been inspired by the recent trend in remote sensing and environmental data acquisition: affordable real-time data management, processing and dissemination are rapidly becoming the norm, but this only applies to streams coming from IoT sensors and mobile devices. Real-time weather data coming from advanced instruments aboard satellites still remains prerogative of government agencies or private spaceflight companies, and is usually available via paid subscriptions. This obstacle can be avoided by using free direct dissemination services, like those offered by NOAA, but those usually require the installation of expensive ground station hardware accompanied by proprietary software. All those problems have been overcome with the development of a custom hardware station and software framework, as it will be explained later in this thesis.

After acquisition, data has to be processed and correlated to generate forecasts and detect unusual phenomena. Human attention works very well when detecting salient patches in an image, and this skill is frequently used to segment areas of interest in satellite images. Due to the relevance of visual saliency in remote imaging and its applicability to many other scientific and psychological research areas, the main part of my work focuses on the study of theoretical saliency models, the research of novel methods for saliency detection in images and the characterization of human attention on sub-object details. As an extension of these, I also investigated the effects of Colour Vision Deficiencies (CVDs) on attention and new techniques to alter image colours depending on saliency values, in order to restore fixation point coherence between CVD affected and normal observers.

1.2 Satellite Remote Sensing

Remote sensing techniques allow us to measure information about an object without touching it. In the last decades remote sensing has been used in various applications such as Earth observation, weather and storm predictive analysis, atmospheric monitoring, climate change, human-environment interactions. Sensors on airborne and satellite platforms have been recording signals from space for many years, giving rise to a huge amount of data. Some data are processed on-board but others are treated and post-processed in ground stations. Signal and image processing are widely applied on data coming from satellites to extract meaningful information for the aforementioned tasks. Satellites and ground stations communicate with each others by using several transceivers and techniques; for instance communication is the largest sector of satellite services. Since the cold war a lot of scientific progress has been made both in navigation and in signal communication and processing. We are interested in communicating with satellites equipped with the Advanced Very High Resolution Radiometer (AVHRR). The AVHRR is a cross-track scanning system allowing the acquisition of signals by using five spectral bands having a resolution of 1.1 km and a frequency of two Earth scans per day. NOAA weather satellites broadcast an APT (Automatic Picture Transmission) and a HRPT (High Resolution Picture Transmission) signal containing a live weather image of the area overflown by the satellite. The scanning lines are oriented perpendicular to the motion of the sensor platform. By using a system that includes a rotating mirror each line is scanned from one side of the sensor to the other. As the platform moves forward over the Earth, successive scans build up a two-dimensional image of the Earth's surface. As said before, the AVHRR scanning system acquires signal by using five spectral bands and, for instance, the visible light image and one of the infrared images are combined in a row vector; the combination is performed using the Automatic Picture Transmission (APT) system.

Despite many scientific progresses over last few years, many scientific issues still affect NOAA satellites: information calibration, physical layer, synchronization, data detection, channel coding. Several scientific experiments are ongoing to overcome those issues, and we focused our efforts to propose a new effective satellite remote sensing tool for signal processing. In greater detail the tool consists of a low cost receiver subsystem for public weather satellites and a signal and image processing module for several tasks such as signal and image enhancement, image reconstruction and cloud detection. Our solution allows to manage data from satellites effectively with low cost components and lightweight computations. As we will show in Section 2 the overall performances are very promising in terms of signal quality and image reconstruction.

1.3 Visual Saliency

The human visual process starts outside the brain with the projection of the light onto the retina. The retina is a thin layer of neural tissue including the rods and the cones that, respectively, allow us to perceive day light and dim light vision. Thanks to the overall architecture of our visual system we are able to receive up to 10 billion bits of information (Li and Gao, 2014). Scientific studies reveal that our cerebral cortex includes 10 billion neurons. It is therefore evident the lack of storage capability of our brain with respect to the huge amount of information coming from our eyes and directed to the cerebral cortex. However, due to the limits of our brain, we cannot simultaneously perform complex analysis on all the input visual information (Li and Gao, 2014). For a given scene, the detection of the most important visual subset occurs as one of the most important task of the Human Visual System (HVS). When a person performs any visual task (watching TV, driving a car) the eyes flick rapidly from place to place to inspect the visual scene. Saccadic eye movements, while observing a scene, are not random: each movement allows the central part of vision (fovea) to fall upon the region of interest of a picture (this is why vision is not uniform across our field of view and acuity decreases with eccentricity) (Snowden et al., 2012).

Attention can be described as the allocation of cognitive resources to information, and it can be divided in five types constituting a hierarchical model (Li and

Gao, 2014):

- Focused attention
- Sustained attention
- Selective attention
- Alternating attention
- Divided attention.

We focused exclusively on selective attention (Sohlberg and Mateer, 1989), that is defined (Li and Gao, 2014) as “the ability to selectively mantain the behavioral or cognitive resource on specific stimuli while ignoring the distracting or competing stimuli”. This is why we conducted experiments and collected results by using object images without visual distractors.

In our work we used an eye tracker to record the gaze path of 24 observers while viewing each image from a subset of an object-based dataset (Viksten et al., 2009a,b) consisting of several images with single objects in the foreground and an homogeneous background color.

We used the eye-tracking data to create a ground truth made of fixation point maps showing where viewers look in the first three seconds of observation.

Scientific studies revealed that the genes for the red and green color receptors are located on the X chromosome, of which males have only one and females have two; this is why color blindness is more widely diffused among males than females. Color vision deficiencies are mainly three, they are caused by protan, deutan, and tritan defects. Deutan color vision deficiencies are by far the most common forms of color blindness. This subtype of red-green color blindness affects about 8% of the male population, mostly in its mild form deuteranomaly (Simunovic, 2010). Red-green color blindness is split into two different types: whereas people affected by protan color blindness are less sensitive to red light, deuteranopia or deuteranomly (the second type of red-green color blindness) is related to sensitiveness on green light. Actually, color vision deficiencies include the followings: protanopia, deuteranopia, tritanopia, protanomalia, deuteranomalia, tritanomalia. The first three are types of dichromacy, which means only two different color receptors (cones) are in the retina instead of three (with normal color vision). The second three (protanomalia, deuteranomalia and tritanomalia) go under the classification of anomalous trichromacy, which means all three different color receptors (cones) are present but one of them is shifted in its peak.

In this work we propose a new method for improving the color perception for people with color vision deficiencies such as protanopia, deuteranopia, protanomalia and deuteranomalia. The main idea behind our work is that saliency maps can be used as crucial information to detect the most important differences between the images as perceived respectively by people with normal and deficient vision systems. We collected eye-tracking observations from people looking at color images with both normal and deficient vision systems. Eye-tracking data have been used both as ground truth and as maps to analyze and highlight the drawbacks of color vision deficient systems with respect to the most important regions according to a normal vision system. To detect automatically the perceptual differences for a given image we first apply a color vision deficiency simulation method and then we extract the saliency maps of both the original image and the version with simulated color vision deficiency. The difference between the saliency maps give us critical information about the regions to be segmented and recolored for the overall image enhancement for CVD people. Once the images have been enhanced they have been used as test for eye-tracking experimental sessions with CVD people to assess the improvement from a perceptual viewpoint. The eye-tracking data we gathered during experimental sessions are made publicly available to be used as a ground-truth under the name of Eye-Tracking of Colour Vision Deficiencies (EToCVD) (Bruno, A. and Gugliuzza, F. and Ardizzone, E. and Giunta, C. and Pirrone, R., 2018).

1.4 Contributions

The main contributions of the work presented in this dissertation are:

- A low cost system for the reception of low resolution images transmitted by NOAA satellite

- A flexible system for the reception of high resolution images transmitted by various weather satellites
- A publicly available object attention dataset
- A saliency method based on color Keypoint Density Maps and optimized for sub-object details
- A multi-scale saliency method based on color Keypoint Density Maps
- A publicly available visual attention dataset of people affected by color vision deficiencies
- A method to enhance images for people affected by color vision deficiencies by using a detailed color mapping of the segmented salient regions of the given image.

1.5 Dissertation Outline

The remainder of the dissertation is organized as follows.

Chapter 2 opens with a short survey of the available techniques for the reception of data from weather satellites and the segmentation of clouds from images. Afterwards, it describes our two hardware/software reception systems and the techniques we used for cloud segmentation. The chapter also reports the experimental results we obtained while developing the low resolution image reception system.

Chapter 3 lists some well-known saliency algorithms, eye-tracking datasets and techniques, and methods for the characterization and mitigation of color vision deficiencies. Afterwards, it describes in detail the eye-tracking sessions we performed and our object attention dataset, Eye-Tracking Through Objects (ETTO); it continues with an explanation of both our saliency methods based on Keypoint Density Maps, and ends with a description of our works on color vision deficiencies: a dataset and a re-coloring method. Every subchapter includes a report of the related experimental results and findings.

Chapter 4 ends the dissertation with the conclusions and some ideas for future research for each topic covered in this thesis.

1.6 Publications

Parts of the work in this thesis have been published in the following referred conference proceedings:

- Edoardo Ardizzone, Alessandro Bruno, and Francesco Gugliuzza. Exploiting Visual Saliency Algorithms for Object-Based Attention: A New Color and Scale-Based Approach. In *International Conference on Image Analysis and Processing*, pages 191–201. Springer, 2017.
- Edoardo Ardizzone, Alessandro Bruno, Francesco Gugliuzza, and Roberto Pirrone. A Low Cost Solution for NOAA Remote Sensing. In *Proceedings of the 7th International Conference on Sensor Networks (SENSORNETS 2018)*, pages 128–134. SCITEPRESS - Science and Technology Publications, Lda. All rights reserved, 2018.

The following paper has been accepted and is pending publication in a book series:

- Francesco Gugliuzza[1], Alessandro Bruno[1], Edoardo Ardizzone, and Roberto Pirrone. An effective Satellite Remote Sensing tool combining hardware and software solutions. In *Communications in Computer and Information Science (CCIS)*, Springer.

The following journal papers have been submitted and are pending review:

- Francesco Gugliuzza[1], Alessandro Bruno[1], Edoardo Ardizzone and Roberto Pirrone. A Multi-scale Color and Keypoint Density-based Approach for Visual Saliency Detection. In *Computer Vision and Image Understanding (CVIU)*, Elsevier.

- Alessandro Bruno, Francesco Gugliuzza, Edoardo Ardizzone, Carlo Giunta and Roberto Pirrone. Image Content Enhancement through Salient Regions Segmentation for People with Color Vision Deficiencies. In *i-Perception*, SAGE Publications.

---

[1] Francesco Gugliuzza and Alessandro Bruno contributed equally to this work.

Chapter 2

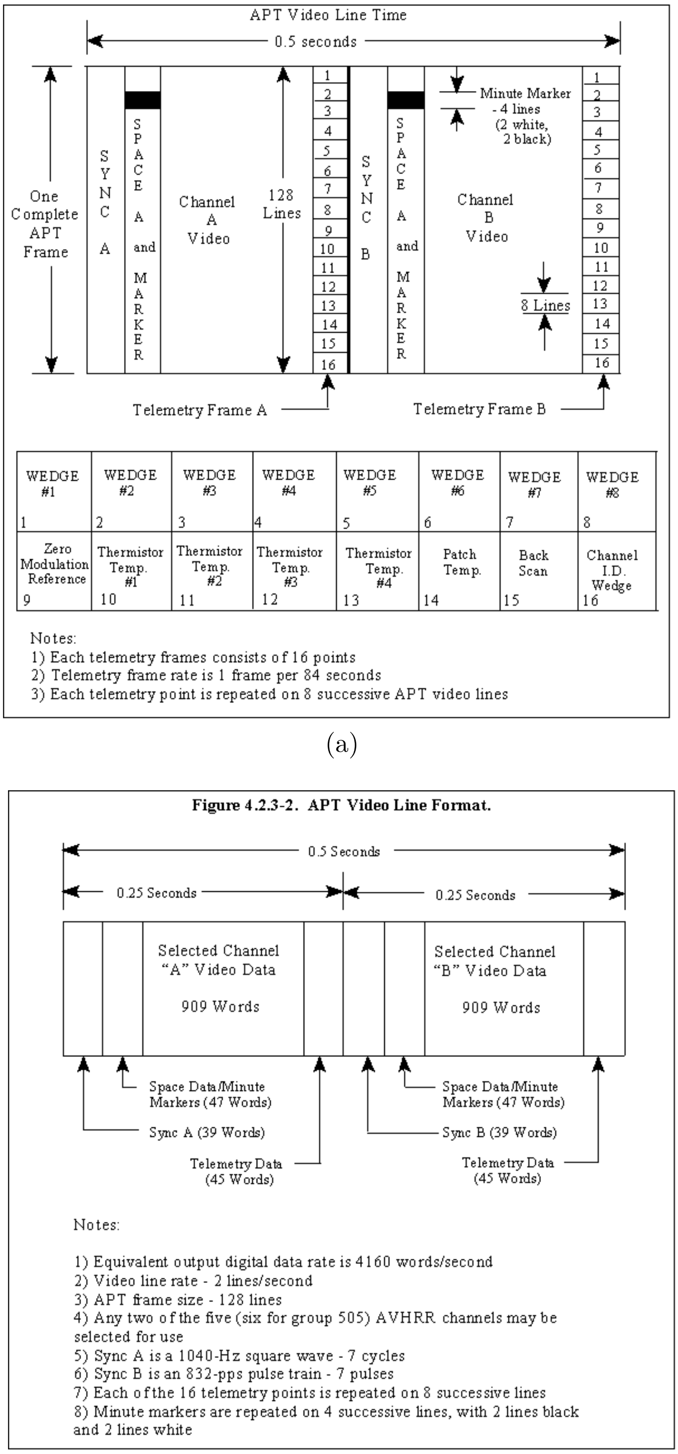
Satellite Remote Sensing

2.1 State of the Art

In this section we give an overview of the widely adopted techniques for the demodulation of signals coming from satellites. Since we perform cloud detection on remote sensing imagery, we give a brief description of techniques of the state of the art for cloud detection.

2.1.1 Satellite Communications and Signal Reception

In this section it is our interest to give the reader a brief description of some aspects related to our research with respect to the satellite communications and signal reception systems. In the last decades satellites have played a critical role interconnecting mankind through complex antennas and receiving systems. Scientific literature shows many technical solutions for antenna elements, sensors, array types, hyperspectral cameras. Depending on the satellite, we can find a low bandwidth communication system (VHF/UHF) or a high bandwidth communication system (more recent solutions have been dealing with communications by using arrays of high bandwidth microwave transponders with the purpose to offer multimedia services (Farserotu and Prasad, 2000)). Scientific community has paid particular attention to reconfigurable antennas and ground receiving station, focusing the efforts towards user demands (Rahmat-Samii and Densmore, 2015). In this respect, several studies have been conducted to assess such aspects as signal prop-



WEDGE #1	WEDGE #2	WEDGE #3	WEDGE #4	WEDGE #5	WEDGE #6	WEDGE #7	WEDGE #8
1	2	3	4	5	6	7	8
Zero Modulation Reference	Thermistor Temp. #1	Thermistor Temp. #2	Thermistor Temp. #3	Thermistor Temp. #4	Patch Temp.	Back Scan	Channel I.D. Wedge
9	10	11	12	13	14	15	16

Notes:
1) Each telemetry frames consists of 16 points
2) Telemetry frame rate is 1 frame per 84 seconds
3) Each telemetry point is repeated on 8 successive APT video lines

(a)

Figure 4.2.3-2. APT Video Line Format.

Notes:

1) Equivalent output digital data rate is 4160 words/second
2) Video line rate - 2 lines/second
3) APT frame size - 128 lines
4) Any two of the five (six for group 505) AVHRR channels may be selected for use
5) Sync A is a 1040-Hz square wave - 7 cycles
6) Sync B is an 832-pps pulse train - 7 pulses
7) Each of the 16 telemetry points is repeated on 8 successive lines
8) Minute markers are repeated on 4 successive lines, with 2 lines black and 2 lines white

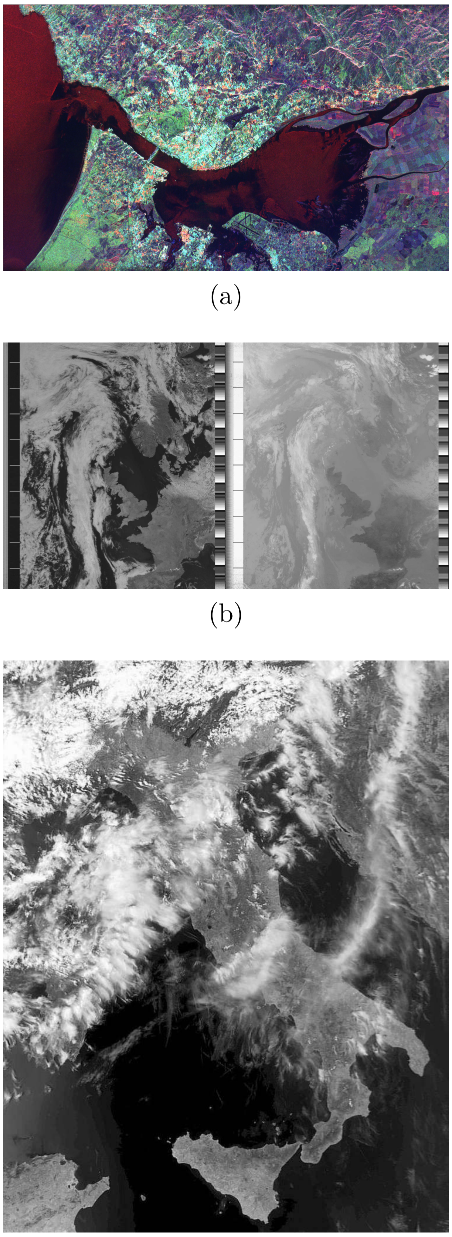
(b)

Figure 2.1: (a) APT frame and (b) row format.

agation, tracking control, higher frequencies, arrays and reflectarrays, mechanical aspects, frequency selective surfaces (Rahmat-Samii and Densmore, 2015).

Modern communication systems allow satellites to receive and transmit simultaneously thousands of signals: radiometers and hyperspectral cameras have been widely used to analyze Earth's visible and infrared radiation from the spacecraft, Synthetic Aperture Radars (SAR) have been used to map surface features and texture, even through dense cloud cover. Due to the aforementioned reasons, we are able to analyze soil/vegetation moisture (to detect anomalies and risks) (Al-Moustafa et al., 2012) or produce 3D models of remote ares of the Earth (up to 1 meter of spatial resolution). Noise and distortions in raw sensor data are usually linked to the sensor nonlinearities. As a general rule they are usually corrected by using classification or regression methods (Camps-Valls et al., 2011).

The APT format was developed around 1960 by U.S. National Aeronautics and Spatial Administration (NASA) as a communication system enabling weather satellites to take pictures over wide areas and transmit them to ground stations on Earth (Barnes and Smallwood, 1982; Wallach, 1997). From that time on, APT has been slightly modified and many stations still keep using analog radio equipment designed in the 1980s or the 1990s. The APT signal is formed by amplitude modulating a 2400 Hz carrier with each line of pixel data obtained by merging two images and calibration data, and then frequency modulating the result; the ease of building a basic receiver composed of a FM demodulator followed by an AM demodulator and an analog-to-digital converter has sparked the interest of researchers, scientists and radio operators. Several aspects such as Doppler effects, free space attenuation (varying with the distance from the satellite) make the reception and demodulation challenging tasks for signals coming from non-geostationary satellites. Furthermore, due to satellite's variable rotation relative to Earth's axis, linearly polarized antennas become unsuitable. For this reason circularly polarized antennas are used to fix the problem, assuring reliable and effective demodulation and synchronization, as well as a good quality output image. Other than demodulation and frequency tracking, image synchronization and calibration are usually performed by means of software solutions rather than hardware systems (Benabadji et al., 2004) because they are non-real time tasks that require some CPU power; it is also important to say that it would be some-what challenging to implement those tasks in hardware (at least for non-experts). Sensor data can then be further corrected to compensate small sensor alignment errors, orbit uncertainty and on-board clock offset, by comparing expected and measured emissivity over land and sea and binning results into a high-resolution grid (Moradi et al., 2013). Over the last years several researchers have dealt with satellite remote sensing tools addressing the issue of affordability of hardware components and software solutions, to accomplish the tasks mentioned in Chapter 1. The authors of (Bosquez et al., 2016) designed a land station receiving through the combination of a software-defined radio (SDR), adaptive antennas on VHF band and FTP servers for the remote analysis of the obtained images. A satellite receiver was designed by Mahmood et al. (Mahmood et al., 2016) with the help of Realtek Software Defined Radio (RTL-SDR), Quadrifilar Helicoidal Antenna (QHA) and Trifilar Balun. QHA and Balun were designed and constructed under a precise and controlled environment. The signals were fed to SDR Sharp (generic SDR software) and were decoded to receive the latest weather image using WXtoImg. In (Uengtrakul and Bunnjaweht, 2014) the authors proposed an experimental kit where a radio signal is received into the computer via the digital TV tuner that operated with the custom driver called RTL-SDR, and then the received signal is processed with a Python signal processing script. In (Sruthi et al., 2013) a low cost alternative to USRP (Universal Software Radio Peripheral) was proposed using RTL-SDR (Realtek Software Defined Radio) which is only used for reception. Furthermore a mixer circuit maps the baseband signal to the band that can be received by RTL-SDR.



(a)

(b)

(c)

Figure 2.2: Examples of (a) a multi-channel SAR image acquired by a Space Shuttle and enhanced by polarimetry (NASA/Jet Propulsion Laboratory, 1999), (b) a two-channel NOAA APT image (Group for Earth Observation, 2003) and (c) one channel of a NOAA HRPT image (Medri, 2005). We are interested in images (b) and (c): in (b) each pixel covers an area of 4 km x 4 km, while in (c) each pixel covers an area of 1.1 km x 1.1 km.

2.1.2 Cloud Detection

Cloud properties such as cloud top height, temperature, infrared radiance have big impact on atmospheric monitoring results (forecasting of meteorological phenomena) (Feidas et al., 2000). Because of the aforementioned reasons cloud detection in remote sensing imagery is a very critical step. It is not straightforward to extract cloudiness mask from a given image from satellite because of several issues with surface properties such as reflectance and emissivity. Since many satellite missions are involved into atmospheric monitoring, achieving good results in terms of detection accuracy is quite important for improving the quality of successive steps of image analysis. Depending on the satellite product we find several features to analyze to detect cloud from an image. In this section we give a brief overview of the state-of-the-art methods focused on cloud detection in satellite imagery. Heidinger et al. in (Heidinger et al., 2012) performed a naive Bayesian approach for cloud detection. This kind of methodology has been adopted for cloud detection on NOAA AVHRR data. In (Alkhatib et al., 2012) the authors used the region growing method for dust cloud segmentation. The authors of (An and Shi, 2015), in order to accomplish the task of cloud detection, proposed an automatic supervised approach based on the scene-learning scheme. In greater detail, they simulate a kind of cubic structural data by considering different image features such as color, statistical information, texture, and structure. In (Ackerman et al., 1997) the authors of the document described the methods used to perform cloud detection in MODIS imagery. Lin et al. (Lin et al., 2015) proposed a radiometric normalization step to perform cloud detection in optical satellite images using invariant pixels. In (Karlsson et al., 2015) the authors examined two probabilistic methods for cloud masking of images from NOAA satellites, obtained with the Advanced Very High Resolution Radiometer. Simpson and Gobat in (Simpson and Gobat, 1996) used AVHRR Split-and-Merge Clustering (ASMC) for cloud detection to overcome the problem of spatially and temporally varying land surface reflectance and emissivity. Split-and-merge clustering allows to segment the scene in its natural groupings and label them as cloud, cloud-free land, uncertain. Gonzàlez et al. in (González et al., 2012) performed cloud classification by using watershed image segmentation, this method has been tested on images from MSG-SEVIRI (Meteosat Second Generation-Spinning Enhanced Visible and Infra-red Imager). The idea behind the method is to segment multispectral images using order-invariant watershed algorithms computed by a multi-dimensional morphological operator. In (Yuan and Hu, 2015) after a super-pixel segmentation step, the authors accomplish the cloud detection task through bag-of-words and object classification methods. Bai et al. (Bai et al., 2016) performed cloud detection with machine learning and multi-feature fusion based on a comparative analysis of typical features such as spectral and texture (Ardizzone et al., 2013b). A supervised approach using a neural network is proposed in (Yhann and Simpson, 1995) for detecting clouds over the ocean using AVHRR data. Other than the aforementioned methods there are deep learning based techniques such as (Shi et al., 2016), in which the authors used Convolutional Neural Networks (CNNs) consisting of four convolutional layers and two fully-connected layers, which can mine the deep features of clouds.

As already mentioned in Chapter 1, visual saliency has been recently used to detect regions of interest in remote sensing images: Zhang et al. (Zhang et al., 2015a) used multi-scale frequency features and pyramids of spatial (color and orientation) features. Techniques based on visual saliency can be used in military applications, by providing the possibility of automatic segmentation of airports (Wang et al., 2013b; Yao et al., 2015; Zhu et al., 2015; Zhang and Zhang, 2017), ships (Wang et al., 2017; Dong et al., 2018) and residential areas (Zhang et al., 2016) from satellite or aerial images.

Deep learning has also been used in this area of research: Diao et al. (Diao et al., 2016) implemented and trained a Deep Belief Network to perform object detection in remote sensing images, while Penatti et al. (Penatti et al., 2015) and Hu et al. (Hu et al., 2015) investigated the possibility of using commonly adopted Convolutional Neural Networks for object detection in remote sensing images.

Some surveys on state-of-the-art cloud detection methods for satellite images, also reporting thorough information on performances and results, can be found in (Chandran and Christy, 2015; Foga et al., 2017; Le Goff et al., 2017).

2.2 Low Resolution Method

Our method focuses on the possibility of directly sampling and processing of the modulated signal entirely in software enabled by recent breakthroughs on Software Defined Radios (SDR) and CPU computational speed. Our objective was achieving good results with low cost SDR hardware like RTL-SDR (Sruthi et al., 2013) (a repurposed DVB-T USB dongle) or LimeSDR: in particular, we had to overcome hardware limitations present in the RTL-SDR receiver: high noise figure and low ADC resolution. We also dealt with the inherent drawbacks caused by frequent tuner saturations. We developed an integrated hardware and software system able to perform the following steps: satellite pass prediction, time scheduling, signal demodulation, image cropping and filtering, and implemented two image processing algorithms to perform cloud detection.

2.2.1 APT Signal

The APT was introduced in 1960s and it shows its age in some respects: it is a mixed modulation signal (AM+FM) carrying a completely analog payload. The standard APT format consists of rows made of 2080 pixels, which are divided in two sub-rows of 909 pixels each belonging to two different sub-images (A and B) and then padded with synchronization and diagnostic information words. Images A and B during daytime are recorded in the visible range and in the infrared range of the electromagnetic spectrum respectively. At night-time, image A is replaced with one recorded at an infrared wavelength different from that of B. Rows can be grouped logically in frames of 128 lines each: a complete frame contains image calibration data and dynamic range references (wedges) (Wallach, 1997). The payload is used to amplitude modulate a 2400 Hz carrier according to (2.1))

$$s(t) = [1 + \mathrm{m}(t)]\ A\ \cos(2\pi f_c t) \tag{2.1}$$

where $\mathrm{m}(t)$ is the modulating signal, A the carrier's amplitude and f_c the carrier's frequency. The amplitude modulated signal is then frequency modulated with a frequency deviation of about 18-20 kHz and transmitted in the 137 MHz band at about 5 W EIRP (36.99 dBm) and 4160 sym/s symbol rate. Right hand circularly polarized (RHCP) antennas transmit the signal, so particular care had to be taken when choosing which antenna type to use in the receiving station.

2.2.2 Hardware

Antenna

NOAA weather satellites require a circularly polarized antenna: the "turnstile" crossed-dipoles design was chosen because of its simplicity and the good performance provided when placed over a ground plane. Ease of construction, durability and suitability of common materials (which can be bought in a hardware store) compensate the drawbacks of the system: the medium gain offered (6 dBi) and many nulls present in such a design (Griffiths, 2014). In particular, the antenna shown here has been built exclusively off PVC pipes, threaded bars as dipole elements, nuts and bolts and 75 Ω generic coaxial cable.

Low Noise Amplifier

The antenna is directly connected to a repurposed old TV amplifier (SIEL .269 01), modified to behave like a wide-bandwidth amplifier. The input variable attenuator and high-pass filter have been removed, and the RF output has been separated from DC power path, allowing power without a bias tee. The amplifier uses two active components:

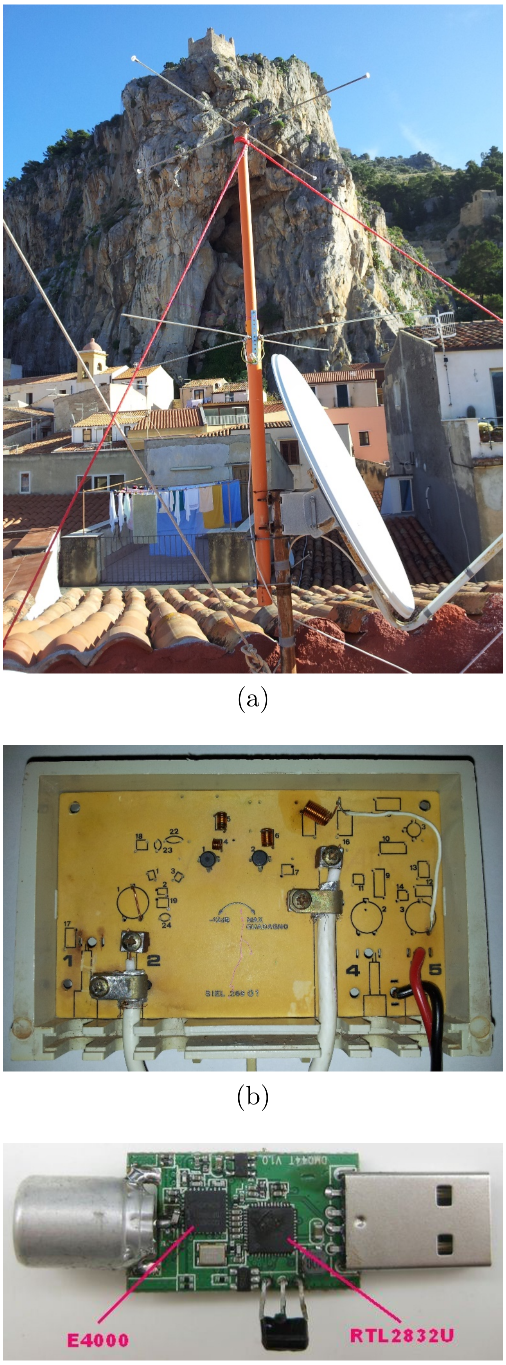
- BFR90A - 16 dB gain, 1.8 dB NF
- BFR91A - 14 dB gain, 1.6 dB NF.

The results are 30 dB gain and $\sim$ 1.83 dB NF (calculated using Friis' formula).

Ignoring noise introduced by amplifier's passive components, and considering a room temperature of 290 K and 40 kHz bandwidth, the noise output of the amplifier can be calculated with formula (2.2).

$$N_{in} + g + NF \tag{2.2}$$

N_{in} is the input noise (-127.95 dBm), g the gain (30 dB) and NF the total noise figure (1.83 dB). The result is a noise output of -96.12 dBm. Input noise has been



(a)

(b)

(c)

Figure 2.3: (a) The turnstile antenna built for 137 MHz reception, (b) the modified TV amplifier, (c) a RTL2832U-based DVB-T dongle with an Elonics E4000 tuner.

calculated using formula (2.3).

$$10\log_{10}(kTB)+30 \tag{2.3}$$

k is the Boltzmann constant, T is absolute temperature and B is signal bandwidth, while +30 has been added to convert results from dBW to dBm. Considering NOAA satellites transmit at 5 W EIRP, when they are at receiving station's zenith the signal power at the receiver's input can be calculated as follows:

$$\text{36.99 dBm - 133.34 dB + 6 dBi - 3 dB + 30 dB = -63.35 dBm} \tag{2.4}$$

133.34 dB is free space loss and 3 dB is the estimated loss due to cable and connectors. Output SNR in best-case conditions is then 32.77 dB, more than enough to ensure good image quality.

Filters

The amplifier's output is connected to a SAW filter (Tai-Saw TA1581A) centered on 137.5 MHz to attenuate out-of-band signals and reduce the severe distortion caused by high-power FM broadcast stations and GSM signals. The TA1581 is a 50 Ω device, but an impedance matching circuit has not been installed because of the low mismatch loss (0.177 dB from antenna to filter and another 0.177 dB from filter to next RF device, which has 75 Ω impedance) compared to filter's 3.5 dB insertion loss. We added a standard ferrite bead near the receiver to shield common-mode interference caused by high-frequency equipment (PC, lab instruments, electronic ballasted lamps, etc.).

SDR Receiver

The SDR hardware is widely available at a very low cost and is built around two chips: a RF tuner and an ADC/COFDM demodulator combo chip. The first used to be an Elonics E4000 (zero-IF), but has been replaced on new models with two superheterodyne alternatives, the Rafael Micro R820T and R820T2. The demodulator chip, a Realtek RTL2832U, is commonly used to demodulate the QPSK or QAM DVB-T signal into a MPEG stream and send it to a PC via an

USB interface, but has a hidden passthrough mode which allows passing the ADC sample stream instead. The samples are internally filtered by a 32 coefficients FIR filter. For this study a dongle containing an E4000 tuner has been used.

2.2.3 Software

The RTL2832 and the tuner require specific initialization and tuning commands, so we decided to rely on a third-party hardware abstraction library coded by the RTL-SDR team (Osmocom, 2012), used by the official Mathworks RTL-SDR support package (Sergienko, 2014). The latter allows easy access to the SDR hardware directly from MATLAB and Simulink environments. As a consequence, this study has been focused on the development of DSP algorithms on Simulink, taking advantage of its model-to-code conversion functionality.

We originally decided to use the very efficient Kyle Keen's rtl_fm (Keen, 2013) FM demodulator, but we later found that it was very easily saturated probably due to some issues in its integer math calculations. Therefore, it was necessary to implement a simple FM demodulation algorithm directly in the Simulink model.

The software we developed consists of the following parts:

- Satellite pass predictor
- Digital down-converter
- FM demodulator
- FIR low-pass filter
- AM demodulators
- Synchronization block
- Pixel dynamic range calibration block
- Model's parameters generator.

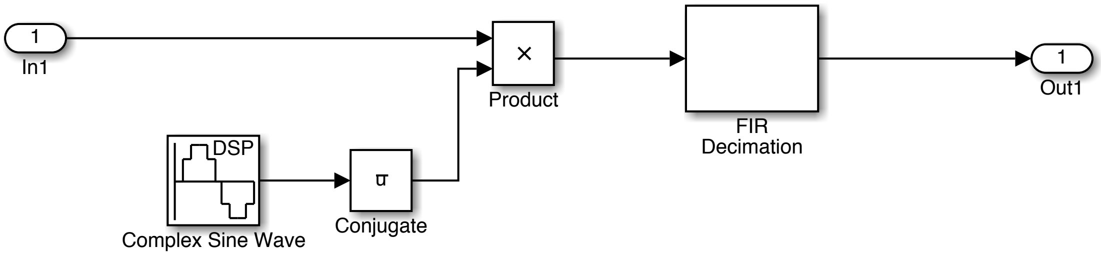


Figure 2.4: Digital down-converter.

Satellite Pass Predictor

Because most weather satellites follow a polar orbit, they periodically pass on the same regions of the Earth. Each pass must be precisely predicted to schedule when to start and stop signal reception and demodulation, so John Magliacane's PREDICT (Magliacane, 2001) software has been used to predict start/end times and maximum elevations of the next satellite passes. In our system this software is called from a script and configured via command line parameters, and its output gets parsed to make it compatible with Microsoft Windows' task scheduler or, in the future, Linux/UNIX's cron scheduler.

RTL-SDR Library for Simulink

The Mathworks RTL-SDR library we used includes all the secondary required files, such as the precompiled low-level hardware abstraction library, sample MATLAB code and a Simulink signal source block.

Digital Down-converter

In order to mitigate the effects of the DC offset present in the Zero-IF tuner, we implemented an offset tuning mechanism: the tuner is set 500 kHz lower than the real signal frequency, and circular frequency shift is performed to shift the satellite's signal back to baseband in software. This mechanism works by multiplying the sampled signal with a complex signal of the type shown in (2.5)

$$e^{-j\Omega_0 n} = e^{\frac{-j2\pi f_0 n}{f_s}} \tag{2.5}$$

where f_0 is the offset frequency and f_s is the sampling frequency.

FM Demodulator

The original implementation of our system included a non-CPU intensive FM demodulator with automatic output amplitude rescaling (to compensate input amplitude fluctuations) (Lyons, 2004), but it was prone to uncontrolled phase shifts in the demoduled output which heavily corrupted the 2400 Hz AM carrier. Therefore a traditional arcotangent-based FM demodulator (Shima, 1995) with amplitude rescaling was used in our work.

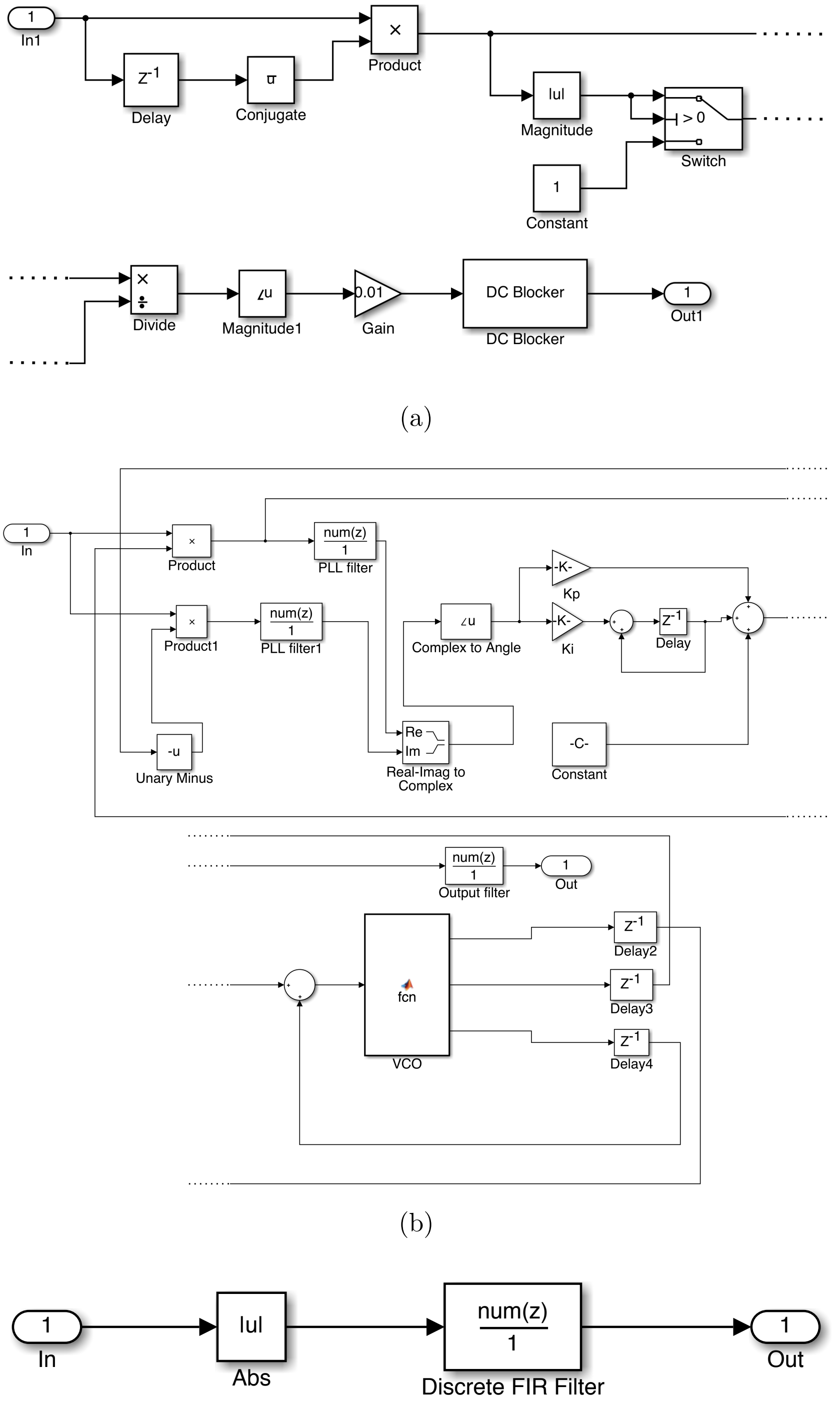
Low-pass FIR Filter

The high frequency noise present in the FM demodulator output is filtered with a linear phase FIR filter designed using a Kaiser window and having a cutoff frequency of 5000 Hz. The latter has been tweaked experimentally by studying receiver and FM demodulator behavior and output frequency content.

AM Demodulators

In our system two AM demodulators are present: the first is based on a discrete-time PLL followed by a low-pass filter with a cutoff of 1500 Hz to attenuate $2\pi f_c$ frequency components. Being a PLL-based receiver, the choice of K_i and K_p coefficients influences heavily its stability, performance and frequency tracking capability, especially when demodulating low SNR signals. We initially considered the delay of the FIR filter when tweaking the coefficients, as in (Wilson et al., 2009), but the demodulator was still plagued by frequent losses of lock and low output SNR. The best results were obtained with the values reported in atpdec's (Leconte, 2003) code.

The second demodulator is based on an envelope detector, similar to analog diode detector demodulators: the absolute value function rectifies the signal and is followed by a linear-phase FIR filter having steep frequency characteristics. This filter greatly attenuates the f_c component while leaving the modulating signal intact and having lower computational complexity than a PLL-based approach. Thanks to the steep characteristics, demodulation is very satisfactory when the



(a)

(b)

(c)

Figure 2.5: (a) FM demodulator, (b) PLL-based AM demodulator, (c) Envelope detector AM demodulator.

input signal has high SNR; when its SNR degrades (as at the start or the end of a satellite pass) it is recommended to switch to the PLL-based demodulator, which is also capable to compensate the Doppler effect to a certain extent.

Synchronization Block

APT format is equipped with two synchronization pulses: horizontal synchronization (sync A) is used as a boundary between video lines and inter-channel synchronization (sync B) is used as a boundary between the two video channels (Wallach, 1997). Our system includes a DSP block that detects synchronization pulses and splits the input signal in image lines composed by 2080 pixels; the image is further decomposed in two sub-images later in the Image Processing code.

The synchronization block requires that the input signal is sampled at a multiple of the symbol rate (4160 symbols/s) and downsamples it by selecting the median of an array of N samples, therefore filtering the signal and rejecting impulse noise.

During the experimental stage we also tested a simple interpolation algorithm for signal sampling, but this solution introduced excessive smoothing to the image and therefore we preferred the median algorithm solution.

Part (a) of Fig. 2.6 down-samples the signal, while (b) correlates previous part's output with a pattern of sync A; synchronization pulses are detected comparing local correlation peaks with maximum correlation in the last 3/4 seconds (c) (time during which one and a half video lines get transmitted). If correlation at discrete instant k-1 is a local peak and is greater than 80% of maximum correlation in said interval, it is considered as a synchronization pulse and a row formed by the last 2080 samples is inserted as a new element at the end of queue (e). Sometimes, however, synchronization pulses are lost or are erroneously detected due to signal fading or noise, so a filtering technique was devised (d): it filters pulses detected less than 2070 samples after last one. If a pulse has not been detected after 2090 samples the line is inserted at the end of the queue anyway and the 2090 samples counter is reset; the 2070 samples counter is not reset not to interfere with subsequent synchronization pulses. Finally, the Rate Transition block (f) outputs two lines of 2080 pixels each per second.

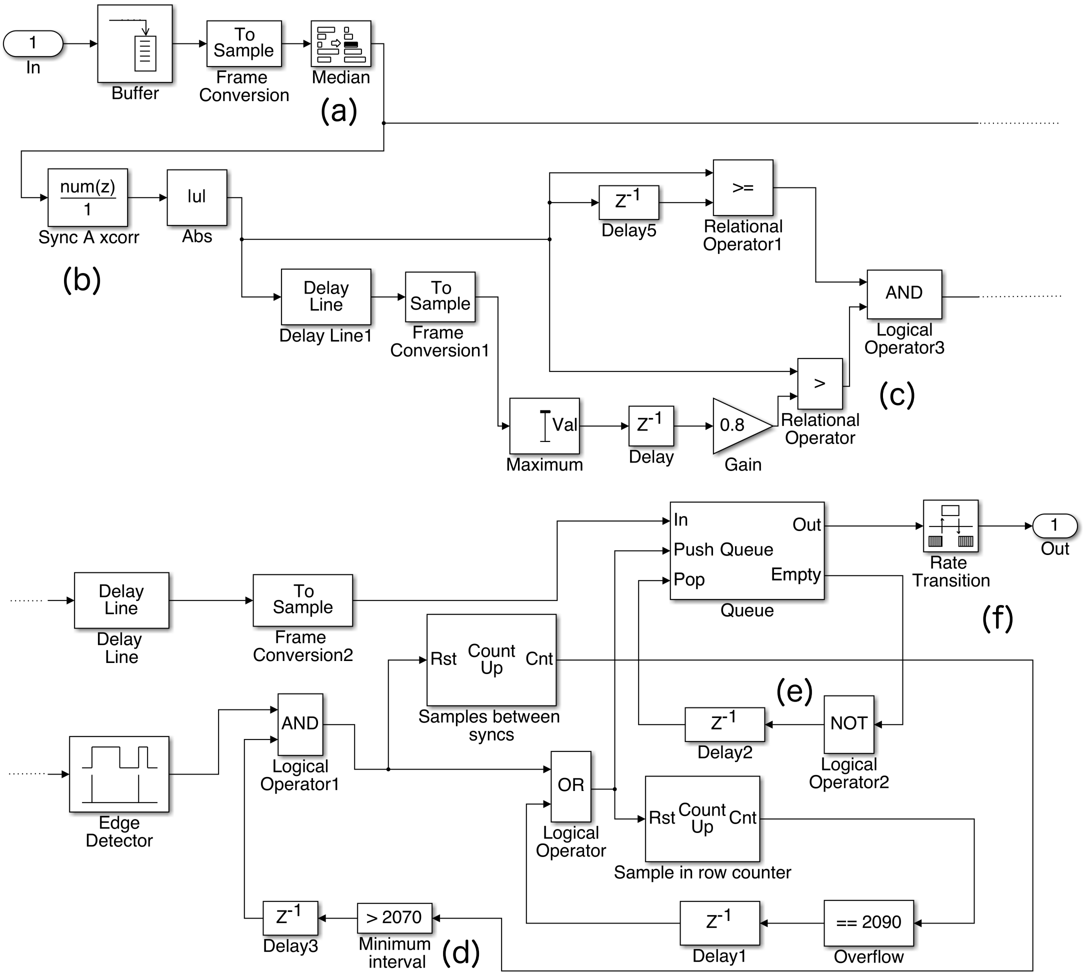


Figure 2.6: Synchronization block.

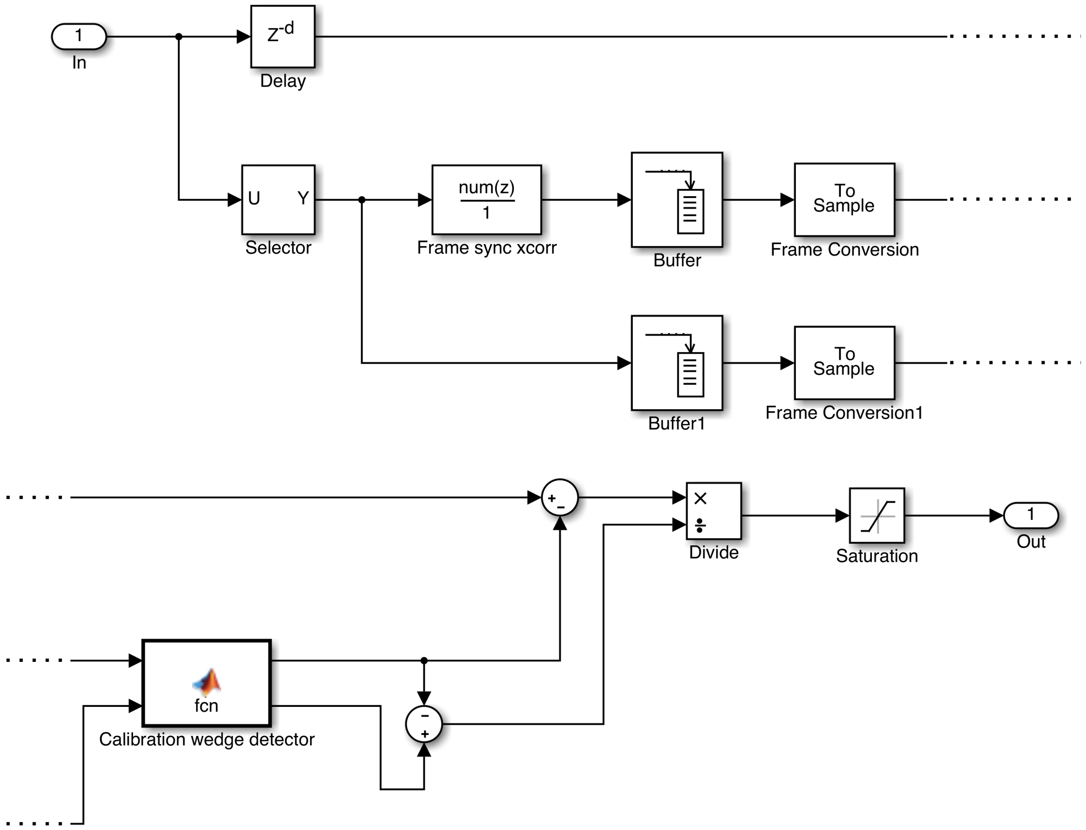


Figure 2.7: Pixel dynamic range calibration block.

Pixel Dynamic Range Calibration Block

Each frame, composed of 128 rows, contains 16 "wedges": the first nine carry data about the image dynamic range (wedges 1-8 represent values from 1/8 to 8/8, while 9 is the zero reference). Our calibration block correlates a column of pixels of the wedge region with an example pattern, and then finds the exact position of each wedge. The values of wedges 8 and 9 are sampled and intensity offset and dynamic range are calculated. Offset is subtracted from pixel values and the result is divided by the detected dynamic range to normalize pixel values to the [0, 1] interval: values outside this range are considered noise or saturation and clipped to 0 or 1.

This block outputs a stream of rows composed of 2080 pixels, which are later

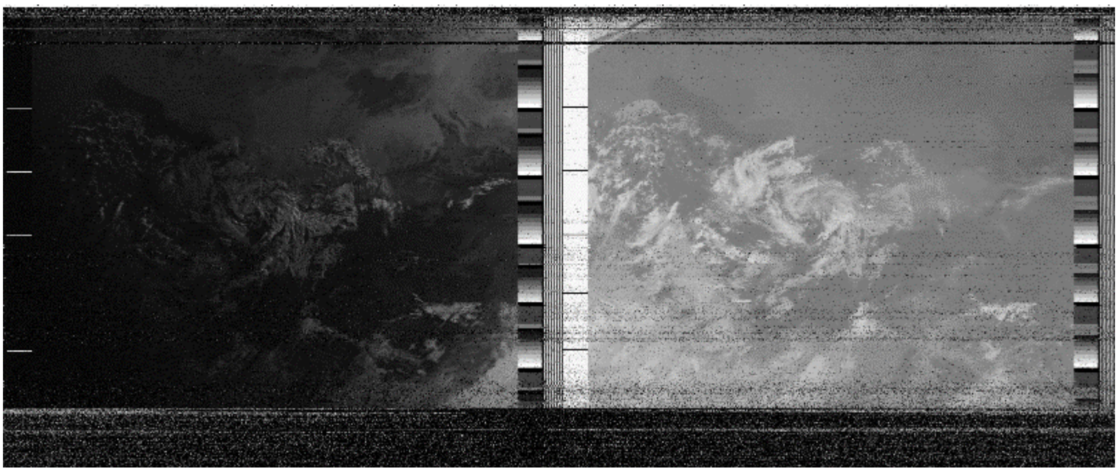


Figure 2.8: AVHRR image after automatic cropping.

processed using Image Processing algorithms.

2.2.4 Experimental Results

Our system generates a 2080x2·T image, where T is running time in seconds. Background white noise or artifacts caused by low signal SNR, fading, interferences and out-of-band intermodulation (Fig. 2.8) are cropped out by an algorithm using pixel variance in calibration wedges to estimate row reception quality. Only rows whose calibration wedges have an entropy below a certain threshold are kept. To further enhance image quality, various Image Processing filters have been applied to the output images and objective metrics (PSNR, RMSE, $\mathrm{SNR}_{\mathrm{rms}}$ and SSIM) have been used to compare the results, which are reported in Table 2.1. The disk filter is shown to be achieving the best results (Fig. 2.9). The results were satisfactory during all the 9 signals reception tests we performed: the PLL-based FM demodulator and processing chain exhibited good performance even under low SNR conditions.

After image denoising we tried two classical approaches to image segmentation, in order to isolate cloud pixels both from land pixels and the rest of the image: the first approach uses the Otsu method (Zhang and Hu, 2008), a global thresholding based on spatial clustering; the one-dimensional version of the method uses only

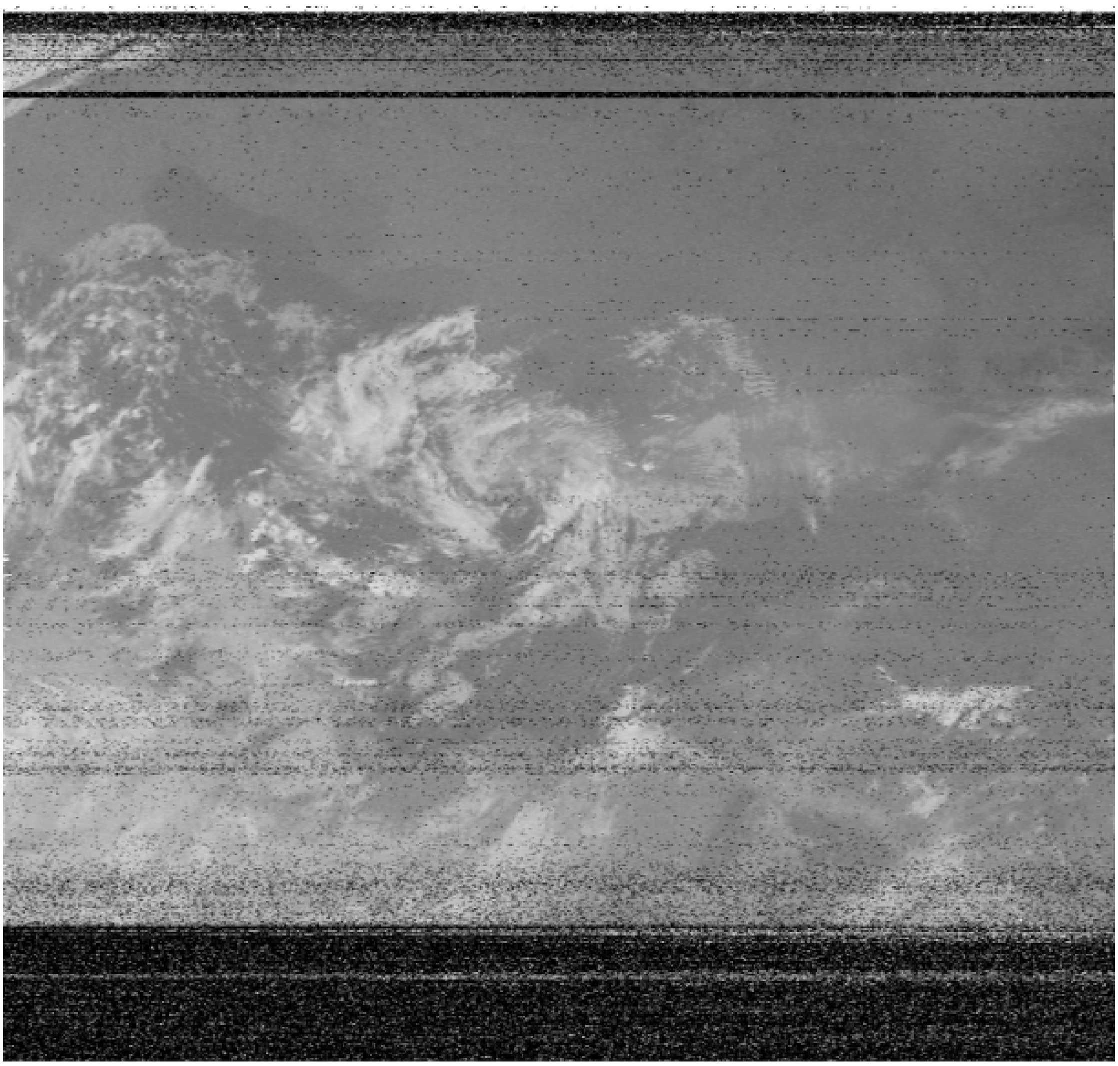
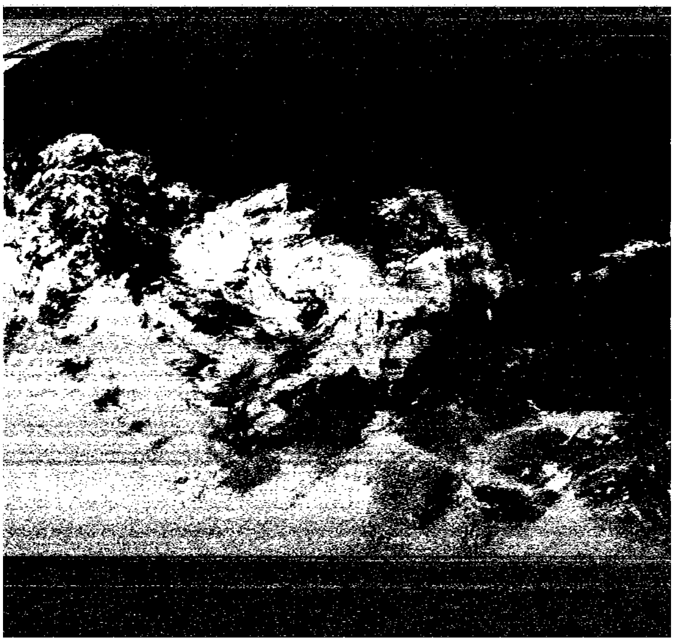
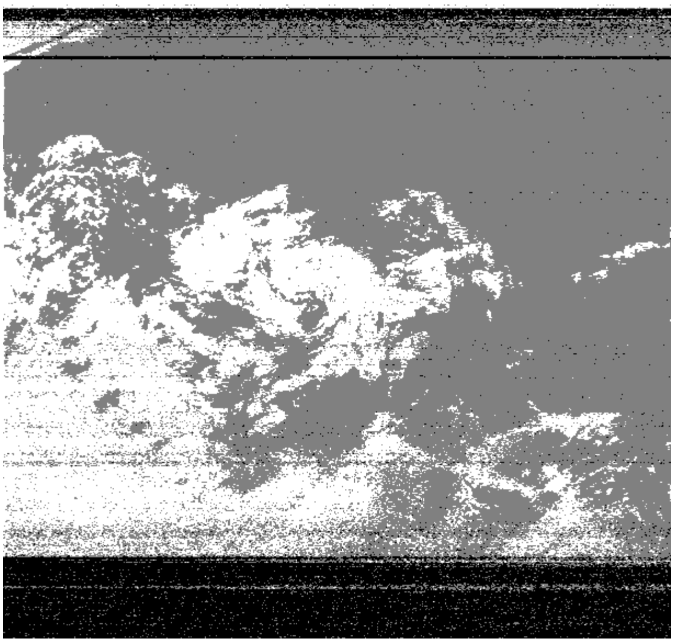


Figure 2.9: Disk filtered remote sensing image from NOAA.



(a)



(b)

Figure 2.10: (a) Cloud segmentation obtained with Otsu method and (b) cloud segmentation achieved using the K-means algorithm ($K = 3$).

Table 2.1: Objective metrics results for image enhancement filtering.

Method	PSNR [dB]	RMSE	$\mathrm{SNR}_{\mathrm{rms}}$	SSIM
None (raw image)	17.788	0.129	4.253	0.557
3x3 arithmetic mean filter	17.981	0.126	4.228	0.606
3x3 geometric mean filter	13.655	0.208	2.488	0.472
3x3 harmonic mean filter	13.295	0.216	2.357	0.449
3x3 contraharmonic mean filter with $Q = 1$	18.023	0.126	4.499	0.625
3x3 gaussian filter	18.485	0.119	4.551	0.609
3x3 disk filter	18.590	0.118	4.570	0.634
Adaptive median filter (min: 3x3 - max: 5x5)	18.444	0.120	4.586	0.622
Adaptive filter 3x3	16.298	0.153	3.550	0.563

the gray level information of the pixel, otherwise the 2D version of the method uses both the gray level information of the pixel and the spatial correlation within the neighbourhood. It is widely used for its simpleness and effectiveness, and is a good solution for images coming from NOAA satellites because it is robust against noise (usually present in APT images); an example of its output is shown in Fig. 2.10a. The second approach is based on K-means segmentation (MacQueen et al., 1967) with $K = 3$, and its results are also satisfactory and are shown in Fig. 2.10b.

2.3 High Resolution Method

Our new high frequency receiving system is designed to receive and demodulate the signals of various meteorological satellites (not only NOAA) transmitting at 1.7 GHz. At this band the signals can have a larger bandwidth and convey data from multiple sensors at higher resolution (e.g. High Resolution Picture Transmission - HRPT (Wallach, 1997)).

Our high frequency system is still in the active development and validation phase, so no high resolution images have been received and demodulated yet.

2.3.1 HRPT Signal

HRPT is encoded and trasmitted very differently than APT. It is a native digital signal modulated using a Phase Shift Keying (PSK) scheme and trasmitted by a circularly polarized antenna at microwave frequencies, around 1700 MHz (Wallach, 1997). The payload contains various high resolution (1.1 km) satellite images at all the wavelengths supported by the radiometer (visible, NIR, MWIR, LWIR etc.); usually it also contains scientific or diagnostic data acquired by other on-board instruments.

There are some variants of HRPT, like Advanced HRPT (AHRPT) and Chinese HRPT (CHRPT): main differences are error correction codes (ECCs), bandwidths and data formats (EUMETSAT, 2013; Zhang et al., 2006). AHRPT is currently in use on EUMETSAT Metop satellites, while CHRPT is used on Chinese FengYun satellites.

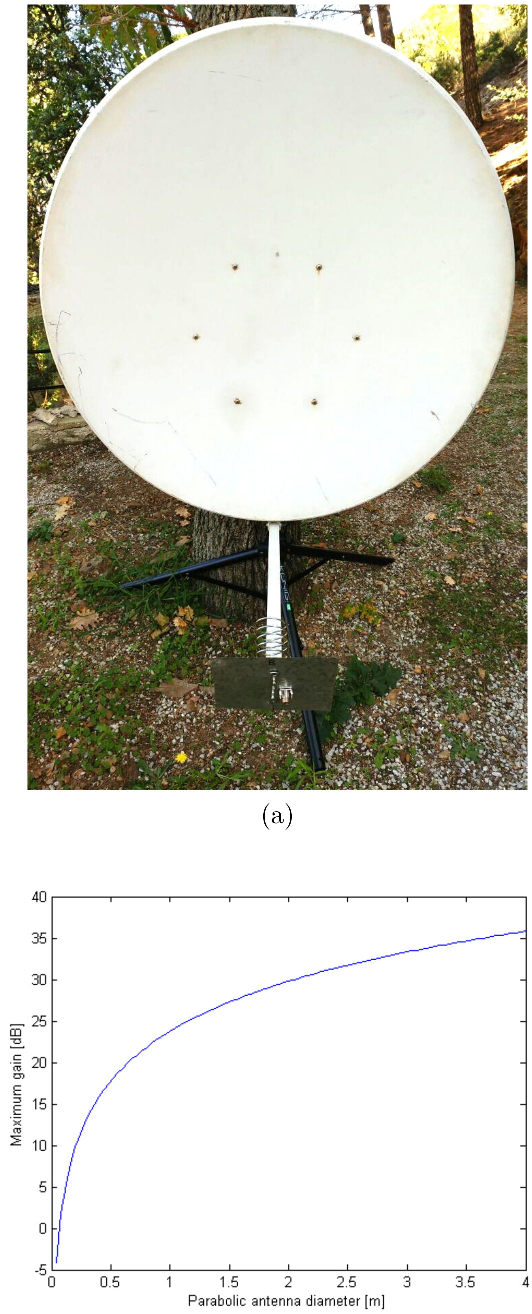
2.3.2 Hardware

Antenna

The system has been designed around an offset parabolic antenna having a diameter of 1.3 m and an efficiency factor of 0.75 (Fig. 2.11a). Because traditional systems have been built using prime-focus dishes having an efficiency of 0.6, our dish can actually have the same performance of a 1.5-1.6 m prime-focus dish, while being lighter and putting less stress on the antenna rotators. As it can be calculated using formula (2.6), where k is the efficiency factor, D the diameter and λ the signal wavelength, our antenna has a theoretical gain of 26 dB at 1.7 GHz (Fig. 2.11b).

$$10 \log_{10} \left[k \left(\frac{\pi D}{\lambda} \right)^2 \right] \tag{2.6}$$

The antenna's structure holds a helical feed designed for a dish having a focus/diameter (f/D) ratio of 0.8 (Fig. 2.12).



(a)

(b)

Figure 2.11: (a) The offset parabolic antenna, (b) theoretical gain vs. diameter of a parabolic antenna having an efficiency of 0.75.

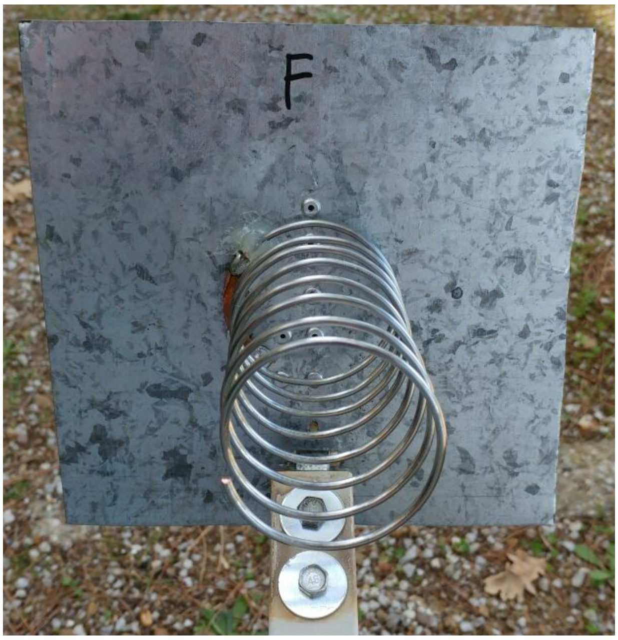


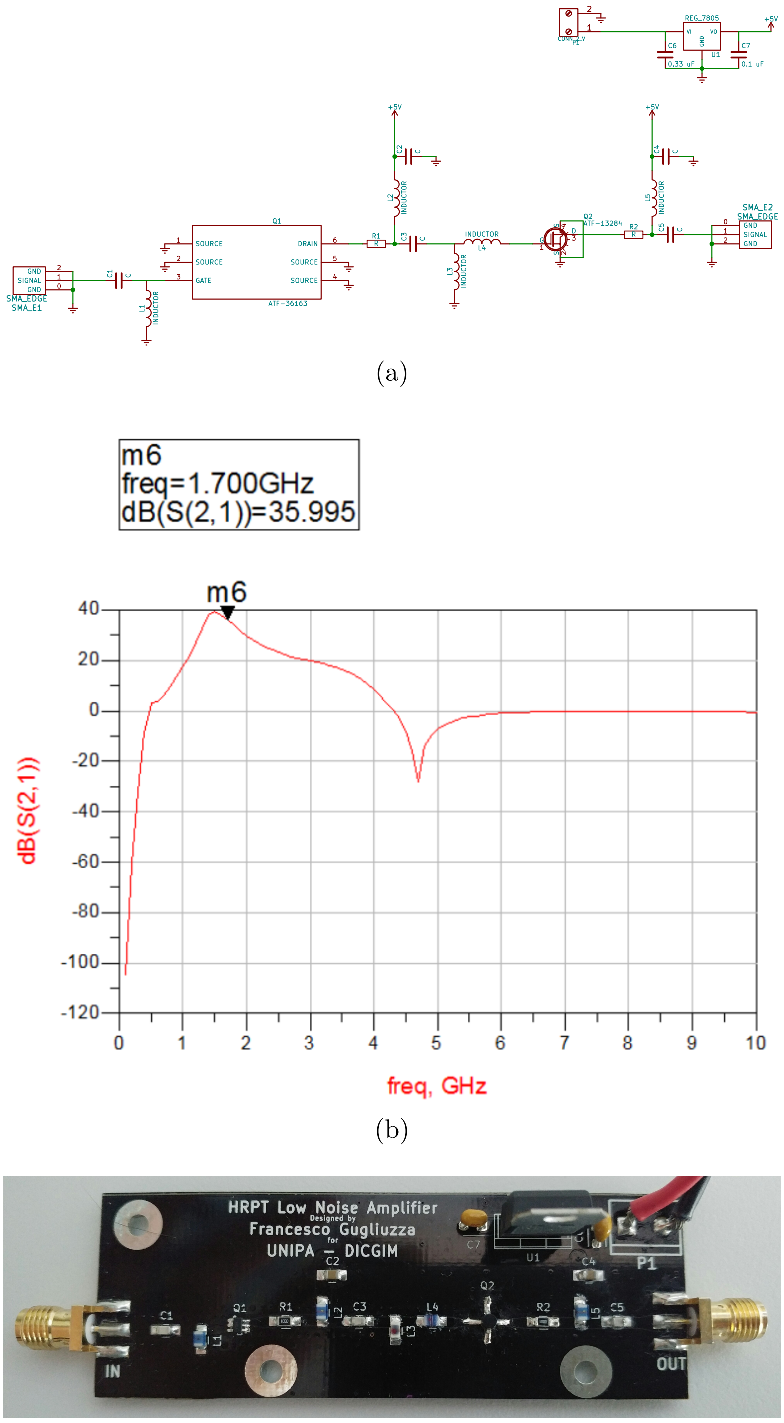
Figure 2.12: The helical feed.

Antenna Rotator

In order to rotate and keep the antenna pointed towards the satellites, a Yaesu G-5500 rotator has been chosen for its durability and reliability, as it has been observed by various amateur radio operators. It can be controlled by the computer via the Yaesu GS-232A interface, which also allows to change azimuth rotation speed to ensure better satellite tracking.

Low Noise Amplifier

In order to amplify the received signal power to an easily demodulable level, we designed and built a dedicated LNA board. It has two stages: the first stage uses an ATF-36163 pHEMT transistor having a very high gain (> 20 dB) and low noise figure (< 0.5 dB) at 1.7 GHz. The second stage further boosts the signal and is based on a more traditional GaAs MESFET, the ATF-13284, which offers a 16 dB gain and 0.6 dB noise figure ad 1.7 GHz. Because those are high frequency and high gain components, PCB design was very critical and we had to take various approaches to avoiding unwanted emissions, instabilities and impedance mismatch.



(a)

(b)

(c)

Figure 2.13: (a) The amplifier's schematic, (b) its gain vs. frequency plot, (c) the populated PCB.

The design has also been optimized for minimum noise figure (0.5 dB) at 1.7 GHz, while also keeping a very high gain (> 35 dB) and being easily reconfigurable by simply changing discrete inductors and capacitors. The circuit schematic can be seen in Fig. 2.13a, while a simulation of its performance is reported in Fig. 2.13b. Finally, the final version of the PCB after it has been populated with components is in Fig. 2.13c.

SDR Transceiver

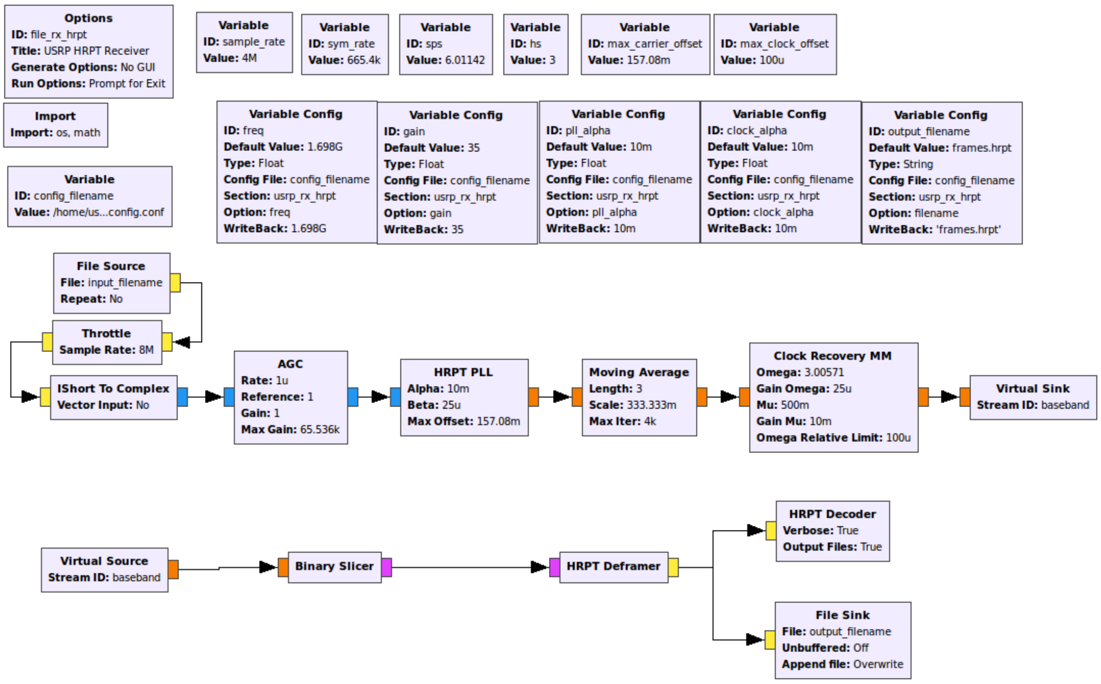
The SDR hardware we chose for our system is the LimeSDR, because it has an extremely wide bandwidth of 61.44 MHz, it is fully open source and it uses the latest Lime Microsystems RF chip, the LMS7002M. The LimeSDR is capable of continuous tuning from 100 kHz to 3.8 GHz and its digital design will allow to offload the computational complexity of the demodulation task from the CPU to the integrated FPGA in the future.

2.3.3 Software

MathWorks Simulink is not optimized for real-time processing of high bandwidth signals, so we chose a different and more appropriate framework: GNU Radio (The GNU Radio Foundation, 2001). GNU Radio offers a flowgraph representation of signal processing algorithms and its blocks are written in Python or C++. Blocks can take advantage of fast and efficient low level math functions via CPU abstractions (VOLK machines (The GNU Radio Foundation, 2015)) or, using third-party libraries, they can directly use GPU acceleration.

Many SDR transceivers and signal processing blocks are already supported in GNU Radio.

Flowgraphs for APT or HRPT signal decoding are already available on the Internet (Csete, 2010; Bülo, 2018) or in the official libraries (The GNU Radio Foundation, 2009) (Fig. 2.14). Our objectives include adding support for other signal types, as Color HRPT (CHRPT), Advanced HRPT (AHRPT), Meteor-M HRPT, High Rate Information Transmission (HRIT), GOES Variable (GVAR) and GOES Rebroadcast (GRB).



Options
ID: file_rx_hrpt
Title: USRP HRPT Receiver
Generate Options: No GUI
Run Options: Prompt for Exit

Variable
ID: sample_rate
Value: 4M

Variable
ID: sym_rate
Value: 665.4k

Variable
ID: sps
Value: 6.01142

Variable
ID: hs
Value: 3

Variable
ID: max_carrier_offset
Value: 157.08m

Variable
ID: max_clock_offset
Value: 100u

Import
Import: os, math

Variable
ID: config_filename
Value: /home/us...config.conf

Variable Config
ID: freq
Default Value: 1.698G
Type: Float
Config File: config_filename
Section: usrp_rx_hrpt
Option: freq
WriteBack: 1.698G

Variable Config
ID: gain
Default Value: 35
Type: Float
Config File: config_filename
Section: usrp_rx_hrpt
Option: gain
WriteBack: 35

Variable Config
ID: pll_alpha
Default Value: 10m
Type: Float
Config File: config_filename
Section: usrp_rx_hrpt
Option: pll_alpha
WriteBack: 10m

Variable Config
ID: clock_alpha
Default Value: 10m
Type: Float
Config File: config_filename
Section: usrp_rx_hrpt
Option: clock_alpha
WriteBack: 10m

Variable Config
ID: output_filename
Default Value: frames.hrpt
Type: String
Config File: config_filename
Section: usrp_rx_hrpt
Option: filename
WriteBack: 'frames.hrpt'

File Source
File: input_filename
Repeat: No

Throttle
Sample Rate: 8M

IShort To Complex
Vector Input: No

AGC
Rate: 1u
Reference: 1
Gain: 1
Max Gain: 65.536k

HRPT PLL
Alpha: 10m
Beta: 25u
Max Offset: 157.08m

Moving Average
Length: 3
Scale: 333.333m
Max Iter: 4k

Clock Recovery MM
Omega: 3.00571
Gain Omega: 25u
Mu: 500m
Gain Mu: 10m
Omega Relative Limit: 100u

Virtual Sink
Stream ID: baseband

Virtual Source
Stream ID: baseband

Binary Slicer

HRPT Deframer

HRPT Decoder
Verbose: True
Output Files: True

File Sink
File: output_filename
Unbuffered: Off
Append file: Overwrite

Figure 2.14: Example flowgraph for HRPT reception in GNU Radio (The GNU Radio Foundation, 2009).

Some parts of the legacy software solution, like the satellite pass predictor, are general purpose and can be reused.

Chapter 3

Visual Saliency

3.1 State of the Art

Models and approaches for visual saliency detection are inspired by human visual system mechanisms. As shown by (Medathati et al., 2016), there is an intimate connection between visual attention and eye movements; for this reason in the last decades, how, why and when we move our eyes is becoming a major topic in scientific research.

Visual attention is mainly guided by two factors: bottom-up factors and top-down factors. Bottom-up factors are stimulus driven, derived from the regions of interest that pop out from the visual scene to our eyes (Snowden et al., 2012). Regions of interest that catch human attention are highly discriminative with respect to the center-surround principle. Visual Saliency aims to imitate the behavior of the human visual system by predicting the fixation points of the most important regions of an image from a perceptual point of view. Visual Saliency is considered to be a multidisciplinary branch of research, laying on the progress achieved by different sciences such as Psychology, Neurobiology, Computer Science, Artificial Intelligence, Medicine. In our work we approach Visual Saliency by using a computational method: we reveal the most salient subset of an image by building the corresponding saliency map. A saliency map is a grayscale map with each pixel falling in the dynamic range [0, 255]. The higher the intensity value, the more salient the location of the visual scene.

As well as visual attention approaches, visual saliency approaches can be grouped by considering the visual feature and the visual attention process involved within the extraction of the saliency map. More in detail, Visual Saliency methods can be grouped in three main approaches: bottom-up, top-down, hybrid.

Visual Saliency bottom-up methods are stimulus-driven, characterized by the so-called “visual pop-out” saliency. In these approaches the exogenous attention is involved with the visual saliency. The center-surround operation (Itti et al., 1998) and graph-based activation maps (Harel et al., 2006) are examples of implemented exogenous attention processes. These methods exploit low level features of the images such as contrast, texture, color, intensity to give rise to saliency maps.

Visual Saliency top-down methods are based on high level visual tasks such as text, object, face detection. In top-down methods the predefined task is given by the object class to be detected (Luo, 2007).

Hybrid methods are conceived to work on two levels: a bottom-up layer allows to extract a noisy saliency map and a top-down layer filters out noisy regions in saliency maps created by the bottom-up layer.

In (Itti et al., 1998) the authors proposed a bottom-up approach based on multi-scale analysis of the image. In greater detail, multi-scale image features are used to create a topographical saliency map, then a dynamical neural network selects the attended locations with respect to the saliency values. The principle of center-surround difference is adopted in (Koch and Ullman, 1987) for the parallel extraction of different feature maps. In (Harel et al., 2006) Harel et al. proposed a saliency method (well known as GBVS) based on a biologically plausible graph-based model: the leading models of visual saliency may be organized into three stages: extraction, activation, normalization. Wang et al. (Wang et al., 2016) surveyed the corresponding literature on the low-level methods for visual saliency.

An effective method (Sun et al., 2016) for visual saliency detection based on multi-scale and multi-channel mean has been proposed by Sun et al. The image is decomposed and reconstructed by using wavelet transform and a bicubic interpolation algorithm is applied to narrow the filtered image in multi-scale. The saliency values are the distances between the narrowed images and the means of their channels. SIFT Density Maps and Keypoint Density Maps have been proposed in our previous works to extract saliency maps and texture scale (Ardizzone et al., 2011, 2013a,b, 2017a). A bottom-up approach is proposed by Aboudib et al. (Aboudib et al., 2015): the authors based their method on Itti and Koch original work in (Itti et al., 1998), but they added a visual angle associated with the image, simulating the fact that viewing a given image from different distances changes the visual angles of the image and, as consequence, it might change the behavior of the Human Visual System (HVS). In (Qian et al., 2013) the authors detect the salient regions of an image by means of contrast feature calculated with various surrounding areas. Otherwise, the authors of (Wang et al., 2013a) incorporated multi-spectral informations for saliency detection. In (Lie et al., 2017) it is proposed a low-resolution saliency estimate based on random color sampling. In (Hou et al., 2013) the non-uniform distribution of the visual information in an image is computed by employing Independent Component Analysis (ICA) to give a measure of the saliency. The authors of (Zhang et al., 2015b) obtained initial saliency seeds by sorting the image boundaries applying a nonlocal anisotropic diffusion equation to the image. The bright and dark regions are usually within a texture and play an important role in terms of perceptually salient attributes, for this reason (Syeda-Mahmood, 1999) specifically analyzed texture features to obtain a saliency measure. A probabilistic approach, associated with the use of graph cut and diffusion equation, for salient object detection has been recently proposed by (Aytekin et al., 2018). Tu et al. (Tu et al., 2017) integrate two saliency maps computed from object proposals and motion-dominated methods, to obtain a spatio-temporal saliency map.

A saliency detection method based on a Kalman filter is proposed in (Roy and Mitra, 2016), inspired by biological phenomena such as the visual surprise and the saccadic eye movement (Bigdely-Shamlo et al., 2008). By adopting the same approach of Itti et al. (Itti et al., 1998), each features channel is individually represented with a generated saliency map by using Kalman filter algorithm, then all of them are combined in a final map.

In top-down approaches (Luo, 2007; Sundstedt et al., 2004), the visual attention process is considered task-dependent, and the observer's expectations and wills analyzing the scene are the reason why a point is fixed rather than others. In (Yang and Yang, 2017) the authors performed saliency detection with a top-down model that jointly learns a Conditional Random Field (CRF) and a visual dictionary.

Kanan et al. (Kanan et al., 2009) adopted the SUN framework to detect the salient regions of an image by using global features and top-down components.

Generally hybrid systems for saliency use the combination of bottom-up and top-down stimuli. In many hybrid approaches (Tsotsos and Rothenstein, 2011; Chen et al., 2003), a top-down layer is used to refine the noisy map extracted from the bottom-up layer. For example the top-down component in (Tsotsos and Rothenstein, 2011) is face detection. Chen et al. (Chen et al., 2003) used a combination of face and text detection and they found the optimal solutions through branch and bound technique. A well known state-of-the-art hybrid approach was proposed by Judd et al. (Judd et al., 2009a) in addition to a database (Judd et al., 2009b) of eye-tracking data from 15 viewers. Low, middle and high-level features of this data have been used to train a model of saliency. Eye-tracking methodology is widely used for tasks such as Human Computer Interaction (Gentile et al., 2016), advertising evaluation (Hervet et al., 2011) and different applications (Duchowski, 2002; Oliveira et al., 2017; Li et al., 2018).

Generally speaking, saliency approaches are based on several properties, features and notions belonging to psychology, computer vision, neuroscience, biology and medicine. In (Mahdi et al., 2017) most of these methods are categorized in seven groups with respect to their computation mechanisms:

- Bayesian models (Zhang et al., 2008)
- Cognitive models (Walther and Koch, 2006)
- Decision theoretic models (Gao and Vasconcelos, 2005)
- Spectral analysis models (Li et al., 2013)
- Graphical models (Salah et al., 2002)
- Information theory models (Wang et al., 2011)
- Learning-based models (supervised learning and unsupervised learning) (Judd et al., 2009a; Vig et al., 2014b).

Yu et al. in (Yu et al., 2016) used a paradigm based on Gestalt grouping cues for object-based saliency detection. In (Chang et al., 2011) the authors proposed a method based on a graphical model of the relationships between saliency and objectness.

The authors of (Toet, 2011) reported a comparative study that evaluates the performances of 13 state-of-the-art saliency models. A new metric is also proposed and compared with previous models. In (Duncan and Sarkar, 2012) the authors gave some formal definitions on three different types of approaches (bottom-up, top-down, hybrid) and an overview on existing methods. Furthermore, the authors offered a description of publicly available datasets and the performance metrics used.

In the last few years, because of their success in recognition and classification tasks, many researchers approached visual saliency by adopting deep learning techniques (LeCun et al., 2015). Deep learning methods allow to extract simple and complex structures in large data sets by using backpropagation algorithm to tune the representation parameters.

For the aforementioned reasons, visual saliency models can be learned by using deep Convolutional Neural Networks (CNNs). For instance, in (Li and Yu, 2016) the authors introduced a neural network architecture, containing a CNN with fully connected layers responsible for feature extraction at different scales. (Das et al., 2017) try to answer the question: do humans and deep networks look at the same regions? For this purpose they conduct qualitative and quantitative comparison of the maps generated by state-of-the-art attention-based models and a task-independent saliency baseline.

The model proposed in (Kruthiventi et al., 2017), called DeepFix, is based on hierarchical learning, and detects semantic features at several levels of abstraction while the large receptive fields in network layers allow to reveal the global context.

In (Cornia et al., 2018) the authors proposed a saliency map method based on deep learning technique achieving good results with respect to the accuracy metrics. In a few words, the saliency prediction architecture proposed in (Cornia et al., 2018) incorporates an Attentive Convolutional Long Short-Term Memory (Attentive ConvLSTM) network, focusing on relevant locations of the image to refine saliency features.

(Johnson-Roberson et al., 2015) used the view frustum and a Hidden Markov Model for calculating saliency. (Zhang et al., 2017) proposed a method for constructing a primitive saliency dictionary: they used representation coefficients and reconstruction errors to count saliency measures. (Chakraborty and Mitra, 2016) detect multiple salient objects in a scene with dense subgraph computation. (Rigas et al., 2015) extract features via local sparse coding on image patches. An overcomplete dictionary is trained using natural images, and a bio-plausible scheme based on the Hamming distance is used to compare patch representations.

A description of the three levels of saliency (Visual Attention Modeling, Salient Object Detection and Salient Object Segmentation), as well as a comparison of the performances of various saliency algorithms and a description the characteristics of the most relevant evaluation datasets can be found in (Furnari et al., 2014). The authors explain why it is necessary to test a saliency algorithm on multiple datasets, and they also show that there is a connection between different saliency levels: eye fixations are suitable for detecting and segmenting salient objects. This leads to the conclusion that saliency algorithms designed to perform those two tasks using bottom-up features could still be improved.

We have recently concentrated our efforts on studying the performance of several visual saliency approaches with respect to the object attention task. We already knew that visual selective attention includes, among others, location-based and object-based attention (Matsukura and Vecera, 2009). Thus, we aimed to analyze visual saliency performance with respect to how attention selects features that are part of an object. In line with this, several experiments conducted over the years demonstrated that observers prefer to make an eye movement to the other end of the same fixated object rather than to an equidistant end of a different object, that is, a preference to make eye shifts within the same object rather than between objects (Theeuwes et al., 2010).

As revealed by findings and researches over last decade (Kim and Lui, 2011; Cheng et al., 2015; Chen et al., 2016a), the relations between low level features of the image and visual attention processes need to be further investigated. For this purpose our investigations addressed in particular the roles played by color and scale features in detecting the visual saliency of a given image. Furthermore, it is observed that most recent saliency detection methods are based on deep learning approaches and achieve high accuracy level in detecting saliency maps but they need high performance systems as a background. In this respect, we investigated several architectures and methods to be compared considering how convenient a solution is in terms of saliency detection performance and software-hardware architecture.

3.1.1 Eye-tracking Technologies and Datasets

Eye-tracking technology offers a direct measure of visual attention by recording two kinds of eye movement, fixations and saccades (Duchowski, 2017). Eye-tracking is widely used in many different task such as the analysis of user behavior in marketing, advertising effectiveness evaluation, neuroscience, human-computer interaction, gaming, medicine, visualization research and other related disciplines (Blascheck et al., 2014). Fixations indicate where the observer actually looks, while saccades are movements between two fixations. Saccades and fixations together form the scanpath. Scanpaths data are used to show which regions of an image catch the observer attention. Visual saliency can be used in several topics and different methods of acquiring fixation points and scanpath data are available in the state of the art.

Eye-tracking technology has improved through the years with the introduction of more accurate instruments and more reliable equipment. Eye-tracking methods can be grouped in four generations (Duchowski, 2017):

- First generation - eye-in-head measurement of the eye consisting of techniques such as scleral contact lens/search coil, electro-oculography
- Second generation - photo and video-oculography
- Third generation - analog video-based combined pupil/corneal reflection
- Fourth generation - digital video-based combined pupil/corneal reflection, augmented by computer vision techniques and digital signal processors (DSPs).

The eye-trackers of the fourth generation that have recently appeared on the market make use of digital optics.

Eye-tracking technology improved its performance in usability, accuracy, and speed by equipping trackers with on-chip Digital Signal Processors (DSPs).

Modern eye-trackers are based on infrared emitters and cameras: the infrared light reflects on the corneas and they appear very bright in the camera image; the position of pupils and reflections can be analyzed and further correction can be applied to account for head position, distortion caused by prescription glasses or small misalignments of the eye-tracker itself. The data is finally converted to x and y positions on the screen and made available through a software API. The devices are usually accompanied by drivers and SDKs that allow integration in video games and other software.

Modern eye-trackers can be grouped in three categories: low, middle and high-end. Low-end devices like the GazePoint GP3, the Tobii EyeX and 4C and the EyeTribe are typically under $1000 and are mainly marketed as gaming peripherals: they have low refresh rate (60 Hz) and require all the calculations to be performed in software on the computer.

Middle-end eye-trackers like the GazePoint GP3 HD and EyeTech VT3 mini are under $10000, are typically aimed at research and incorporate higher FPS cameras and on-board processing. The latter is required to offload mathematical computations at high refresh rates (up to 200 Hz) and avoid saturating the communication bus.

High-end eye-trackers like the Tobii TX 300, the EyeLink 1000 and the Smart Eye Pro are over $10000 and boast features like higher refresh rates (300 Hz and beyond), compensation for large head movements and extreme accuracy and precision. In this range we can also find some glasses-mounted trackers (e.g. the Tobii Eye Tracking Glasses 2), which allow integration with head mounted displays or data acquisition in situations where a user is not sitting behind a screen but, instead, he/she is roaming around in an environment.

The recent progresses in eye-tracking technology allow scientific researchers to collect a large quantity of eye gaze data for different purposes.

Jiang et al. collected (Bappy et al., 2016) a dataset named EyeCrowd by recording the eye movements of 16 subjects watching images with various levels of crowd. Judd et al. (Judd et al., 2009a) created the MIT dataset, composed of 1003 landscape and portrait images and the corresponding fixation point maps, collected during a free viewing session.

Some recent datasets such as (Lin et al., 2014) and (Shen and Zhao, 2014) are respectively focused on domains like visual saliency in low resolution images and web pages.

The authors of (Ramanathan et al., 2010) focused on the eye-tracking of face, portrait, nude, action, affect-variant groups giving rise to a dataset composed of 99 fixation point maps. The aforementioned datasets are collected during free viewing session, that is, the subjects were not assigned any specific visual task when watching. On the other hand, task driven eye-tracking data are found in such works as (Ehinger et al., 2009).

A very popular fixation data set containing 2000 images from 20 different categories has been proposed in (Borji and Itti, 2015).

In our experiments we followed the same data gathering protocol used by Torralba et al. (Judd et al., 2009a).

3.1.2 Image Content Enhancement through Salient Regions Segmentation for People with Color Vision Deficiencies

Biological science focused on molecular genetics underlying color vision (Neitz and Neitz, 2011). Machado et al. (Machado et al., 2009) simulated color vision by using a physiologically-based model and handling normal color vision and color vision deficiencies such as anomalous trichromacy and dichromacy in a unified way.

A lot of water passed under the bridge since Shinobu Ishihara proposed the series of plates as test tool for color-blindness consisting of 38 isochromatic plates (Ishihara, 1960): the plates form an easy method of establishing the diagnosis and distinguishing cases of red-green deficiencies. The plates are held 75 cm from the subject and tilted so that the plane of the paper is at right angle to the line of vision. Since then, several models have been proposed as tool to detect color vision deficiencies.

The Farnsworth-Munsell 100-Hue (FM100) tests (Farnsworth, 1957) is a standardized measure of chromatic discrimination, based on colored cap-sorting, which has been widely used in both adults and children. During FM100 test it is asked to order the shown color plates in the correct order, any misplacement can be related to a sort of color vision deficiency (Vingrys and Cole, 1983). The RGB

Anomaloscope color blindness test consists of two different lamps with different lights to be matched and it is a well known and accurate tool to classify color blindness. It was developed by a German ophthalmologist more than 100 years ago and it is still being used internationally to check color vision deficiencies and specific subtypes (Lakowski, 1969).

A pseudoisochromatic color plate test called Color Vision Testing Made Easy (CVTMET) has been proposed by Cotter et al. (Cotter et al., 1999). It was designed for all age groups, it uses the identification of simple shapes and objects to detect red-green color deficiencies. Bimler et al. quantified variations in color spaces with respect to sex differences (Bimler et al., 2004). A web application written in JavaScript has been presented in (Gambino et al., 2016) by Gambino et al. implementing a digital Ishihara-like test for pre-school aged children. In (Chen et al., 2016b) Chen et al. delivered a color-blindness image (CBI) in order to deliver direct and effective information to dichromats by transforming color-blindness images into the pattern-highlighted image. Transform is made by means of color component analysis, pattern attention as well as thresholding. The experiments confirmed the improvements of processing steps on CBI by means of Ishihara test plates.

A lot of progress has been made in the last decades on simulating color vision deficient systems (Meyer and Greenberg, 1988; Brettel et al., 1997; Kondo, 1990; Walraven and Alferdinck, 1997; Ichikawa et al., 2003, 2004). Machado and Oliveira (Machado et al., 2009) proposed a method aimed at simulating the loss of chromatic contrast transforming the RGB image into a orthogonal dichromatic color space. Tajima and Komine (Tajima and Komine, 2015) developed a method based on visual saliency for quantifying and visualizing information loss and gain resulting from individual differences in spectral sensitivity. An algorithm (Rasche et al., 2005) that transforms color to grayscale preserving image detail by maintaining distance ratios during the reduction process is proposed by Rasche et al.

Some methods of the state the art focused on the enhancement of colored regions with respect to visual attention perspective. Huang et al. (Huang et al., 2009) approach is based on grouping the colors on CIE L*a*b* space through a Gaussian Mixture model.

eyePilot (Perception Data Inc., 2006) is a fairly useful technique developed to assist color blind people in understanding and working with color-coded information. Jeong et al. (Jeong et al., 2012) proposed an image re-coloring method based on color clustering with an information preserving property for color-blind people.

We focused our attention on how effective the enhancement of salient regions of an image is with dichromatic vision systems. We used visual saliency like a tool to detect the most important differences between normal and color vision deficient systems.

3.2 Eye-Tracking Through Objects (ETTO) Dataset

An experimental session has been conducted to record the eye-tracking data and therefore to validate our research. We focused our experimental session on the analysis of the object attention, 5 objects have been selected from the Object Pose Estimation Database (OPED) (Viksten et al., 2009a,b) to be observed by 24 subjects (males and females between 21 and 34 years old). The database consists of several images with single objects in the foreground and an homogeneous background color, but any dataset with a single main object (target) and a limited number of distractors in each image would have been appropriate as well. We selected 19 views having a 130° fixed vertical angle and an horizontal angle ranging from 0° to 180° in 10° increments. The resulting 95 images have been processed to filter the noise that could grab human attention. Then a padding operation has been applied to the images to fill the 22 inch screen at 1920x1080 resolution. During the experimental session the subject was placed at approximately 70 cm from the screen (Fig. 3.1). To ensure an accurate recording of the eye movements, the Tobii EyeX (running at 60 Hz sampling rate) has been calibrated to each user. Each image was shown for 3 s while the eye-tracker was capturing all user's saccades and fixations, then a neutral gray screen was shown for 1 s before showing the next image. This protocol allowed us to keep the results consistent to those of other works in literature (Judd et al., 2009a).

The data acquired from the eye-tracker consist of three arrays of the same length: an array with the coordinates of the fixation points of the left eye, an array with the coordinates of the fixation points of the right eye, the sampling time values are stored in a third array.

We computed the coordinates of the fixation points by averaging the coordinates of the first two arrays, that is, the coordinates of both eyes. The results have been converted to screen coordinates, then the real fixation points map (at full resolution) have been computed by adding one to the pixel value, each time a subject looked at that pixel.

The data collected with the eye-tracker needed to be in a continuous form to be compared with saliency maps. Therefore the real fixation maps have been convolved with a Gaussian filter and then normalized to [0, 1] (Fig. 3.2).

Our Eye-Tracking Through Objects (ETTO) dataset is publicly available (Ardizzone et al., 2017b) and contains the the coordinates of each fixation and with all the corresponding real fixation point maps.

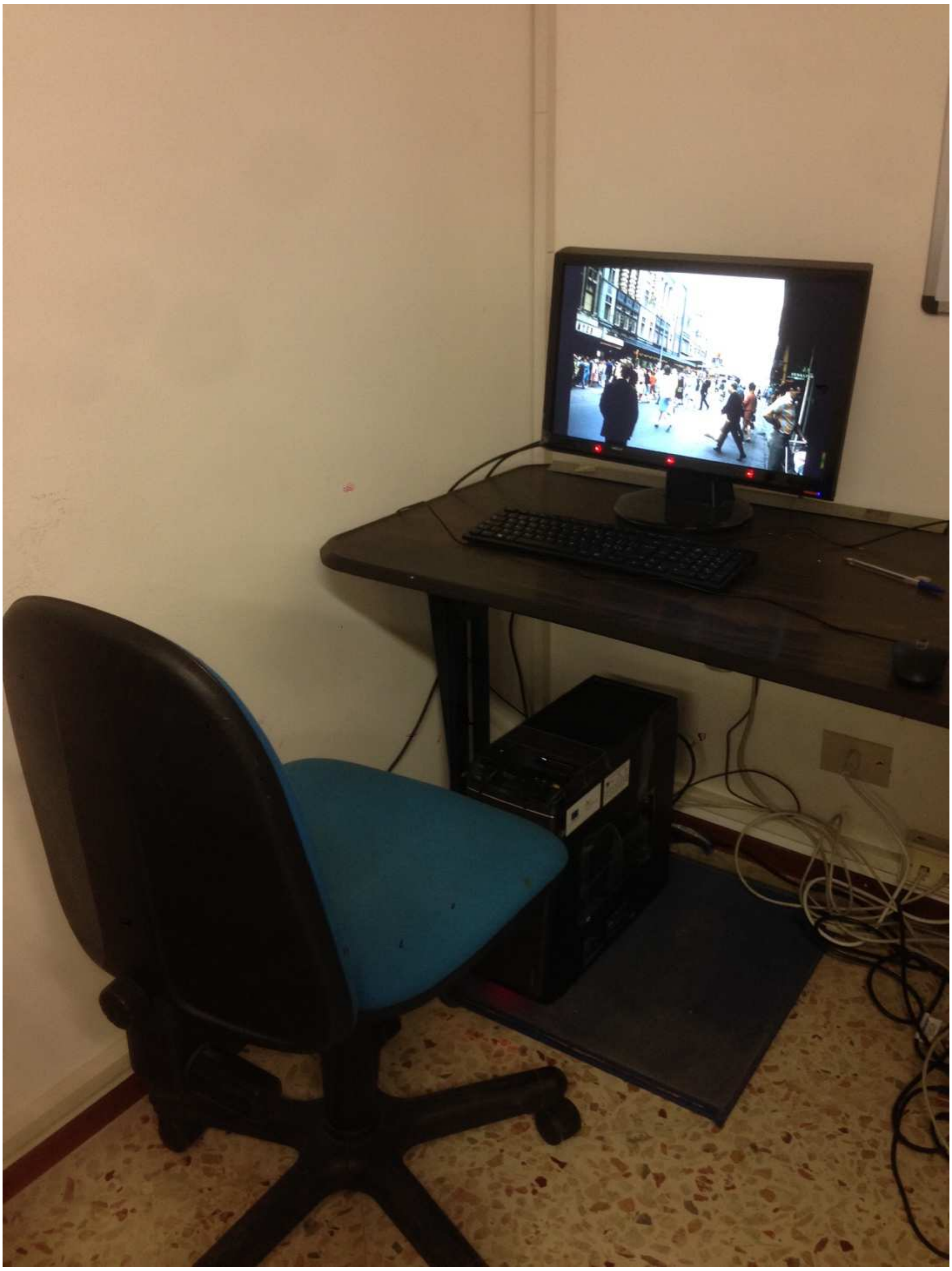
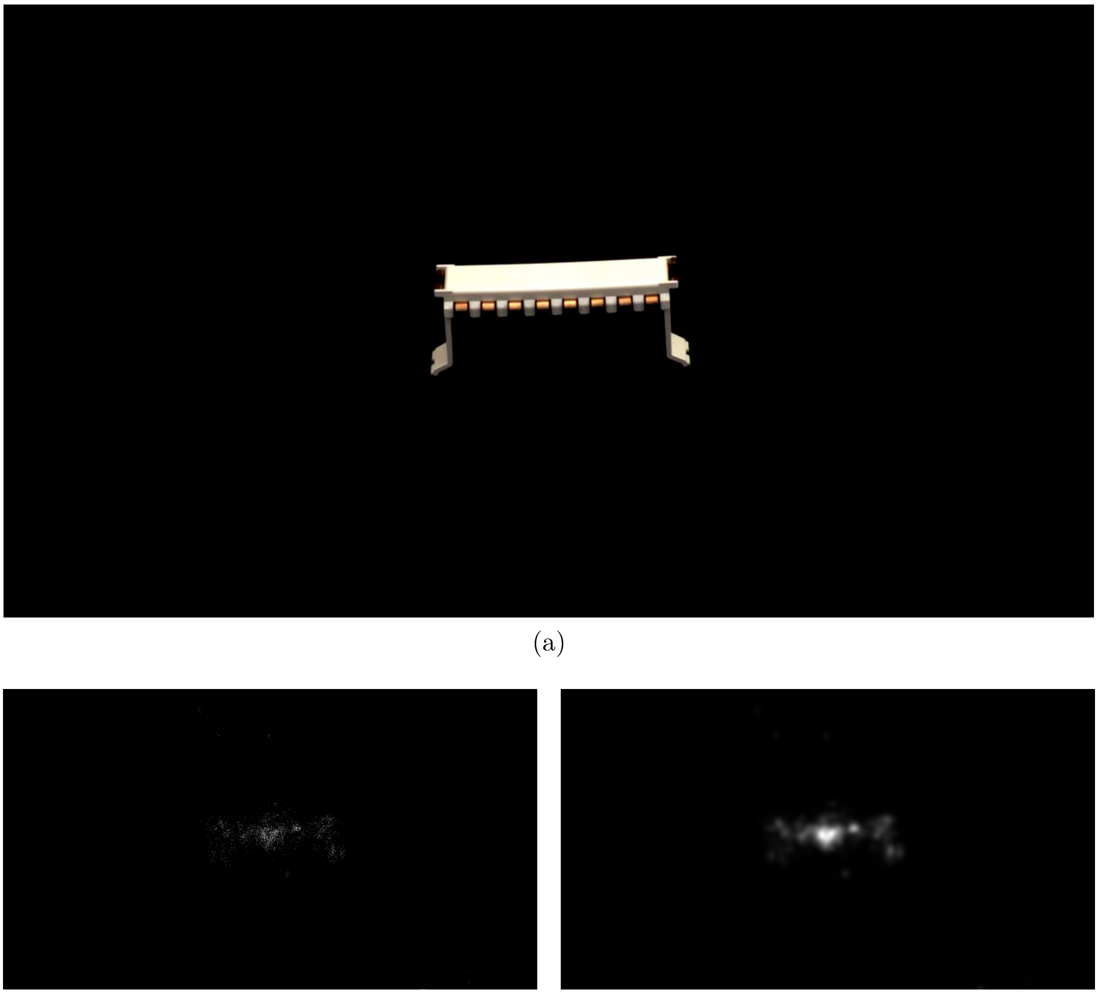


Figure 3.1: The setup used for eye fixation data acquisition.



(a)

(b)

(c)

Figure 3.2: (a) An image from the OPED and its fixation map (b) before and (c) after Gaussian blurring.

3.3 Color SIFT Density Map Method

We aimed to improve the method developed in 2011 (Ardizzone et al., 2011) by adding chroma information to the saliency map generation algorithm based on SIFT (Lowe, 2004) Density Maps (SDMs). A SDM is built by counting the number of detected SIFT keypoints inside a sliding window of size k x k centered on each pixel of the image. To obtain a valid saliency map, the SDM is further processed by taking the absolute difference of each pixel with the most frequent value (mode) of the map, rescaling the values to [0, 1] and blurring the result with an average filter which has a window size that is half of that used to build the map (k).

Color-based saliency has been implemented in two ways by harnessing the power of HSV and CIE L*a*b* color spaces. We early found that the optimal SDM window size equation we used in (Ardizzone et al., 2011):

$$k = 2^{\lfloor \log_2 \left(\frac{\min(M,\ N)}{4} \right) \rfloor} \tag{3.1}$$

where M and N represent the horizontal and vertical resolution of the image, is unsuitable in object attention because it is calculated on entire image size, while the object only takes a small central portion of it, causing excessive loss of detail in the generated saliency maps. We overcame the problem by first taking the mean of the dimensions of the object bounding boxes in all images used during the data acquisition phase, then applying (3.1) to the calculated values.

3.3.1 HSV Color Space Saliency

In HSV an image is expressed using cylindrical coordinates, where hue is an angular dimension that goes from 0° to 360° and then back to 0°, while saturation and value are linear dimensions. 8-bit RGB images can be easily converted to HSV by projecting the RGB cube on a chromaticity plane in such a way that an hexagon is formed:

$$\begin{aligned}
C_{max} &= \max(R, G, B)\\
C_{min} &= \min(R, G, B)\\
\Delta &= C_{max} - C_{min}\\
H &= \begin{cases} 0 \text{ if } C_{max} = 0 & \\ \left(60 \times \frac{G-B}{\Delta} + 360\right) \bmod 360 & \text{if } R = C_{max}\\ 60 \times \frac{B-R}{\Delta} + 120 & \text{if } G = C_{max}\\ 60 \times \frac{R-G}{\Delta} + 240 & \text{if } B = C_{max} \end{cases}\\
S &= \frac{\Delta}{C_{max}}\\
V &= \frac{C_{max}}{255}
\end{aligned} \tag{3.2}$$

We convert hue and saturation from polar coordinates to cartesian coordinates[1], in order to eliminate the discontinuity around zero in hue values:

$$\begin{aligned} X &= S \circ \cos(H)\\ Y &= S \circ \sin(H) \end{aligned} \tag{3.3}$$

then we rescale the X, Y, V channels to the [0, 1] range for convenience of processing and separately calculate statistically processed SDMs. The three maps are combined into the final saliency map using the following formula:

$$SM_{HSV} = \frac{1}{3}(SM_H + SM_S + SM_V) \tag{3.4}$$

where SM_H, SM_S and SM_V are the saliency maps of hue, saturation and value, respectively.

3.3.2 CIE L*a*b* Color Space Saliency

HSV space still shows some shortcomings, namely hue and saturation channels are dominated by noise when brightness is low; furthermore, it is not biologically

---

[1] Not to be confused with CIE XYZ.

(a) (b)

Figure 3.3: (a) An image from the OPED and (b) its SIFT saliency map calculated in HSV space.

(a) (b)

Figure 3.4: (a) An image from the OPED and (b) its SIFT saliency map calculated in L*a*b* space.

inspired, and does not model the HVS color opponent process (Engel et al., 1997). Therefore, we decided to implement SDM calculation also in the CIE L*a*b* space, which is perceptually uniform and designed with color opponency in mind (Sharma, 2002).

The processing steps are essentially the same as the previous method: RGB → L*a*b* conversion (D65 illuminant used as reference), channel range rescaling, SDM calculation, statistical processing and fusion. Coordinate transformation has been omitted because this color space does not have mathematical discontinuities.

3.3.3 Experimental Results

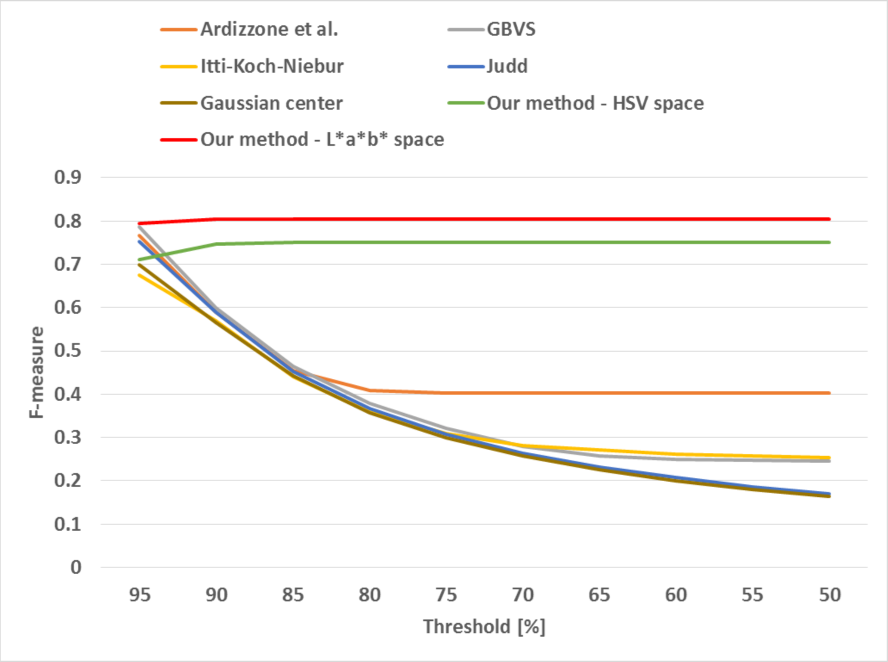
We generated saliency maps using various methods, as our legacy work (Ardizzone et al., 2011), Itti-Koch-Niebur (Itti et al., 1998), GBVS (Harel et al., 2006), Judd (Judd et al., 2009a), our two new color-based methods and a fixed centered Gaussian distribution as a baseline (Judd et al., 2009a). We ran tests on our 95 image dataset and its related fixation point and fixation map database, on an Intel Core i7-4770 computer with 4 cores (8 threads) and 16 GB of RAM. For the calculation of GBVS and Itti-Koch-Niebur saliency maps the GBVS Toolbox (Harel, 2012) has been used, as it includes an enhanced implementation of Itti's algorithm; Judd saliency maps were instead generated running Judd's code (Judd et al., 2009b) with its original trained parameters. We binarized saliency and fixation maps at various percentiles (Judd et al., 2009a; Ardizzone et al., 2011) (between 0.95 and 0.5) and evaluated the performance of our method in terms of F-measure values:

$$\begin{gathered} P = \frac{|M_D \cap M_R|}{|M_D|}; R = \frac{|M_D \cap M_R|}{|M_R|} \\ F_1 = 2\frac{P \cdot R}{P + R} \end{gathered} \tag{3.5}$$

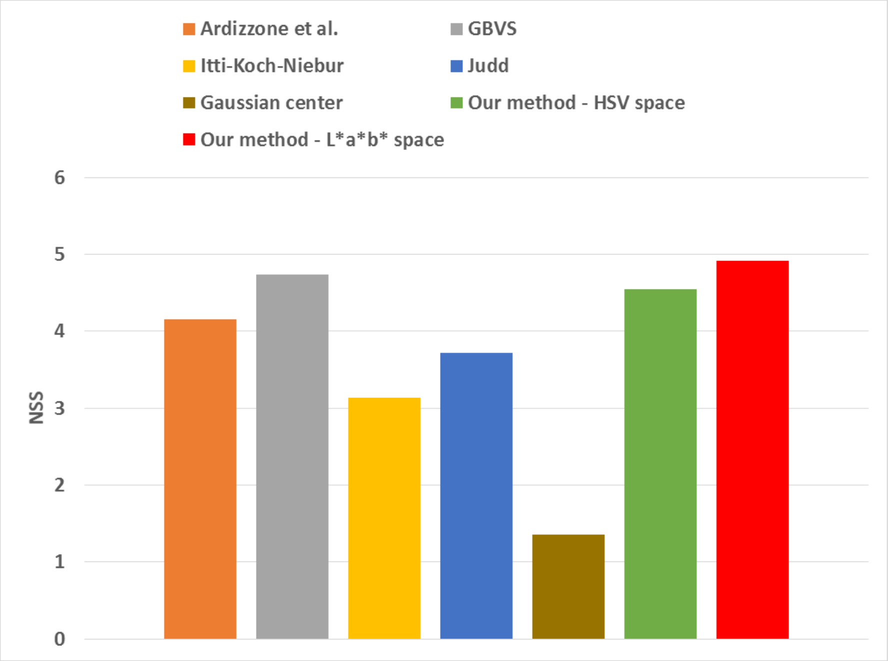
where M_D is the binary version of the detected saliency map, while M_R is the binary version of the reference fixation map. We also calculated Normalized Scanpath Saliency (NSS) values, which is a well balanced, binarization-independent metric (Bylinskii et al., 2018).

From Fig. 3.5, we note an opposite trend with respect to natural image saliency model performances reported in other works: in object attention, as saliency threshold reduces the F-measure tends to reduce as well, instead of increasing. The performances of both our models, instead, increase slightly with threshold until they reach a plateau at 90% saliency levels. Our CIE L*a*b*-based method always gets best results in both metrics, while the HSV-based method underperforms at high saliency levels with respect to GBVS and our previous work.

The execution time required for calculating a HSV or L*a*b* saliency map is about 12 s for a 1920x1080 image.



(a)



(b)

Figure 3.5: Performance graphs of various saliency models in terms of (a) F-measure vs. threshold and (b) NSS values.

3.4 Multi-scale Color SURF Keypoint Density Map Method

3.4.1 Keypoint Density Map

In this subsection we introduce the concept of Keypoint Density Map (KDM), which is the evolution of the SIFT Density Map. The following explanation has been adapted from (Ardizzone et al., 2013b), the paper in which the KDM has been first introduced.

A KDM is a representation of the density of a distribution of the key-points in an image and can give essential information about the regularity of its structure.

Let us consider a M x N image I, with an extremely regular distribution of keypoints. The average number of pixels per keypoint, N_P, is:

$$N_P = \frac{M \cdot N}{n} \tag{3.6}$$

where n is the number of image keypoints. Given

$$s_1 = \sqrt{N_P} \tag{3.7}$$

a squared area of size s_1 x s_1 will include only one keypoint, regardless of its position within the image. In general, a squared area of size s_k x s_k, where

$$s_k = \sqrt{k \cdot N_P} \tag{3.8}$$

(as a matter of facts, all the values s_k are rounded to their nearest integer values) *on the average* will include k keypoints. Due to construction, actually the window will include exactly k keypoints only when $\sqrt{k}$ is integer. Otherwise the number of keypoints included in the window will depend on its position (e.g. a window of size s_2, that is the square root of $2N_p$, may include 1, 2 or 4 keypoints of the image.

A Keypoint Density Map KDM$_k$ is built by counting the number of keypoints into a sliding window of size s_k. Each point in the KDM$_k$ indicates the number of keypoints falling into a squared area of size s_k, centered in the corresponding point of the image. $k_{\max}$ is limited by the image size:

$$\begin{aligned} k_{\max} &= \frac{1}{N_P} \cdot s_{k_{\max}}^2 \\ s_{k_{\max}} &= \frac{1}{2} \min(M, N) \end{aligned} \tag{3.9}$$

3.4.2 Multi-scale KDMs and KDM Improvements

Here we extend our previous works (Ardizzone et al., 2011, 2017a) in which we investigated the relationships between local keypoint detectors (Ardizzone et al., 2013b) and the behavior of the human visual system when a person looks at an image.

In this section we report our findings and studies with some methodological novelties based on color features and a new multi-scale inspection. Furthermore, we report a complete comparison study with respect to the best and most popular saliency detection methods and techniques.

Here we would like to highlight some noteworthy aspects: we develop computational methods for the extraction of saliency based on pattern recognition and image processing techniques. As mentioned above, we already proposed (Ardizzone et al., 2013b) the Keypoint Density Maps (KDMs) for texture scale detection in grayscale images which show regular and near regular textures.

We created a small sample dataset composed of 100 images for test and experimental purposes, using 20 randomly selected images from each of these datasets (Ardizzone et al., 2017a; Judd et al., 2009a; Borji and Itti, 2015; Bylinskii et al., 2015a; Jiang et al., 2014).

As already said in Section 3.3, to improve the effectiveness of our algorithm, in the past we added color information (Ardizzone et al., 2017a) from two color spaces such as HSV and CIE L*a*b* and experimental evidence showed several drawbacks in HSV space, thus we did not notice the improvements we expected by adding HSV color information. Scientists noticed that HSV is not biologically inspired and does not take into account the color opponent process (Engel et al., 1997).

The higher performance achieved through the use of CIE L*a*b* space prompted us to undertake a deeper investigation on perceptually uniform color spaces (CIE

L*a*b*, CIE L*u*v*).

In this work we compare the performances of two local keypoint detectors, SIFT and SURF, when extracting saliency maps by means of KDMs method and adopting two perceptually uniform color spaces (CIE L*a*b* and CIE L*u*v*). We also improved KDM generation and smoothing by adding a circular mask to the sliding window and a Gaussian filter in place of the arithmetic mean filter. A multi-scale approach was implemented, as it will be explained below.

Local keypoints are extracted on each channel of the color space, then KDMs are computed as a measurement of the spatial distribution of keypoints over the entire image.

The KDM algorithm needs a scale factor to work: this parameter is function of the size of the image (Ardizzone et al., 2017a).

The keypoint spatial distribution gives a measure of the behavior of texture inside the image. Texture allows us to read the behavior of the image with respect to several features such as contrast, scale, orientation, edges, object boundaries. According to our method, we extract salient regions by emphasizing rare events in textured regions. One of the basic concepts is that the spatial distribution of keypoints inside an image allows to describe texture variations all over the image. In our previous works we applied this principle by using SIFT keypoints only on grayscale images (the standard SIFT algorithm (Lowe, 2004) is suitable for grayscale images only). Encouraging results prompted us to approach a method including color information because of its importance in cognitive terms. As explained above, in (Ardizzone et al., 2017a) we tried to use HSV space color and it did not allow us to achieve the expected results in terms of effectiveness and accuracy, then we concentrated our efforts on the extraction of saliency maps by using KDMs in all the three channels of two perceptually uniform color spaces: CIE L*a*b* and CIE L*u*v*.

We measured the performance of our legacy algorithm on the sample dataset in both CIE L*a*b* and CIE L*u*v* spaces, with and without circular masking and Gaussian filtering. As shown in Fig. 3.8, the CIE L*u*v* method with circular masking and Gaussian filtering had the best results, therefore subsequent improvements were only tested on this method.

We extended the KDMs used to detect the texture scale in regular and near regular textures, to the extraction of the saliency maps of an image by applying them to all of the three color channels. Once the keypoints are found in each color channel, four KDMs having different k are calculated for each channel by selecting a "main" scale factor, as in equation (3.1); the other three k are selected by simply subtracting 1, 2, 3 to the main k. If $k = 1$ is reached before calculating all the four maps, no more maps are calculated for that channel.

For each KDM, a saliency map is calculated by taking the difference between the KDM and its mode value (3.10)

$$\mathrm{SM}_{i,j}(x,y) = \left|\mathrm{KDM}_{i,j}(x,y) - \mathrm{MV}\left(\mathrm{KDM}_{i,j}\right)\right| \tag{3.10}$$

where i is the index of the color channel and j is the scale index.

Then, inspired by the fact that human gaze patterns follow a normal distribution in the natural viewing condition (Sugano and Bulling, 2015) we combined the saliency maps by using polynomial regression in the 2D subspace. In order, the twelve saliency maps $(3 \cdot 4)$ are passed to a third degree polynomial without mixed terms to perform pixel-by-pixel multi-channel and multi-scale fusion, as shown in (3.11)

$$\mathrm{SM}(x,y) = a_0 + \sum_{i=1}^{3}\sum_{j=1}^{4}\sum_{l=1}^{3} a_{i,j,l}\left[\mathrm{SM}_{i,j}(x,y)\right]^l \tag{3.11}$$

where i and j represent the same indices as in (3.10).

The result is finally multiplied with a center bias map (Fig. 3.6).

The coefficients of the polynomial in (3.11) have been trained by performing a regression on the reference fixation maps of the sample dataset. A 10-fold cross-validation (9 training subsets - 1 validation subset) has been used and the best of the 10 models has been selected; during training, the multi-scale maps have been pre-weighted by a center bias map and scaled to the same spatial resolution (512x512) before passing them as an input to the polynomial. The algorithm has been implemented in MathWorks MATLAB and the polyfitn toolbox (D'Errico, 2012) has been used to perform the regression and evaluate the polynomial.

Our objective was, therefore, to analyze how color information and a multi-scale approach help to improve the effectiveness of our method with respect to different ground truth datasets composed of real fixation maps (Ardizzone et al.,

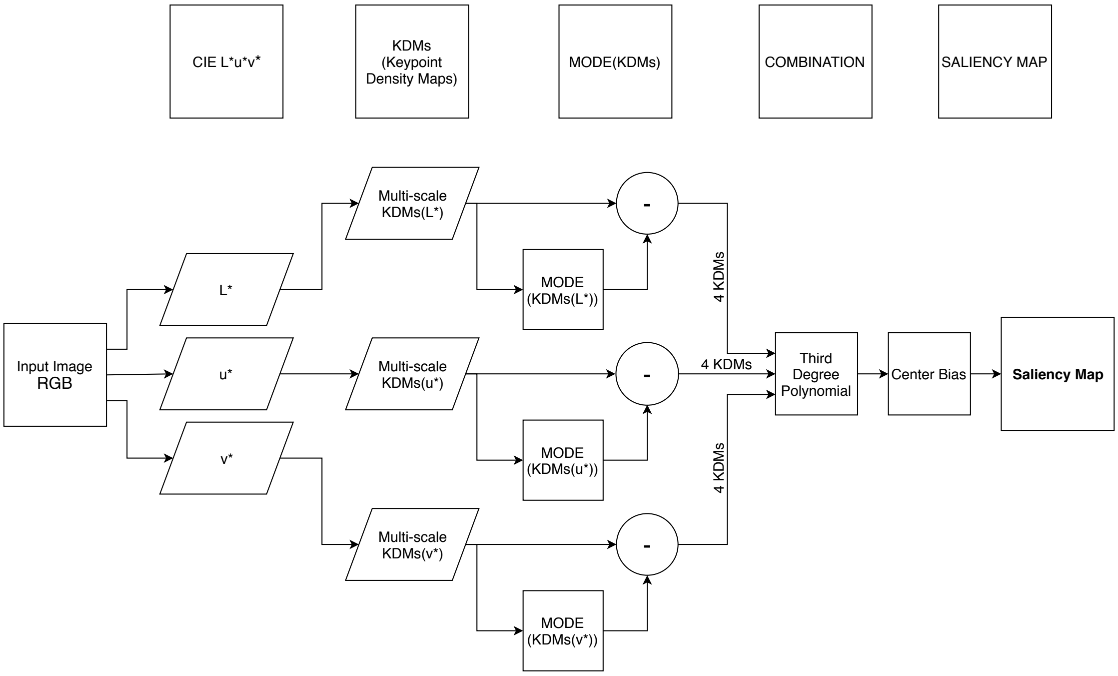


Figure 3.6: The block diagram of the overall steps of saliency map extraction for a given RGB input image.

2017a; Judd et al., 2009a; Borji and Itti, 2015; Bylinskii et al., 2015a; Jiang et al., 2014).

Furthermore, we investigated the effectiveness of our method focusing on two processes: visual object attention and free viewing. To do this, we tested our approach on several datasets, some of which are dedicated to the processes cited above.

We remark that the spatial distribution of keypoints inside an image can be used in texture variation description. The levels of roughness of both fine and coarse regions can be very different (in a fine region we will find a larger number of keypoints than in coarse regions), so we used keypoint density maps, to find various texture events and to identify the most salient regions.

We also added a center bias module to tackle the human-gaze center bias, that is, a simple Gaussian blob centered in the middle, as described in (Judd et al., 2009a). As we will show in the Experimental Results section, the performances of our methods definitely improve by adding the center bias module.


(a) (b)

Figure 3.7: (a) An image from the OPED and (b) its saliency map calculated in L*u*v* space using the SURF keypoint detector, multi-scale map generation, polynomial fusion and center bias weighting.

3.4.3 Experimental Results

In this section we compare our saliency methods both to classical algorithms and more recent methods based on deep learning, using our own dataset and generic image datasets. Our aim is to show the performance of our methods with various image types contained in generic datasets, as emotional photos, landscapes, hand-drawn sketches, etc. The attention datasets we used for this study, all accompanied by eye-tracked fixation locations, are the following:

- ETTO (Eye-Tracking Through Objects - our dataset) (Ardizzone et al., 2017a)
- MIT1003 (Judd et al., 2009a) - it is a collection of eye-tracking data of 15 viewers (males and females between the ages of 18 and 35 years) on 1003 images collected from Flickr and LabelMe that have a maximum resolution of 1024x768
- MIT CAT2000 (Borji and Itti, 2015) - it is a collection of eye-tracking data of 120 viewers (40 male, 80 female with an overall mean age of 20.1 years) on a set of 2000 images from 20 different categories; each image has been viewed by 24 different observers. Publicly available data is limited to the fixation maps of only 18 observers

- FIGRIM Fixation Dataset (Bylinskii et al., 2015a) - it is a collection of eye-tracking data of 40 observers (16 male, 24 female with an overall mean age of 21.2 years) on a set of 2787 images from 21 different categories; each image has been viewed on average by 16 different observers
- EyeCrowd (Jiang et al., 2014) - it is a collection of eye-tracking data of 16 observers (10 male and 6 female, between the ages of 20 and 30 years) on a set of 500 natural crowd images with a diverse range of crowd densities.

The saliency algorithms used for comparison are the following:

- Our new multi-scale CIE L*u*v* SURF-based method
- Our legacy CIE L*a*b* SIFT-based method (Ardizzone et al., 2017a)
- Itti-Koch-Niebur (Itti et al., 1998)
- GBVS (Harel et al., 2006)
- ConvLSTM-based Saliency Attentive Model with a VGG-16 network (SAM-VGG) (Cornia et al., 2018)
- ConvLSTM-based Saliency Attentive Model with a ResNet-50 network (SAM-ResNet) (Cornia et al., 2018)
- Ensembles of Deep Networks (eDN) (Vig et al., 2014b).

We also report the performance of a fixed centered Gaussian distribution as a baseline and choose Normalized Scanpath Saliency (NSS) as our comparation metric as before, because it is well balanced and binarization-independent (Li and Gao, 2014); we report other metrics (Bylinskii et al., 2018) to better show the improvements and the effectiveness of our method.

The Itti-Koch-Niebur algorithm we chose for comparison is an enhanced version released with the GBVS Toolbox (Harel, 2012), which also contains the official GBVS code. The reference implementations of ConvLSTM-based and eDN models have been downloaded respectively from (Cornia et al., 2017) and (Vig et al., 2014a). All the algorithms reported above have been used with their default parameters. As explained in the previous section, the main scale factor k in our method is computed according to equation (3.1), while the coefficients of the polynomial in equation (3.11) are learned through the regression.

All the saliency maps have been calculated on an Intel Core i7-4770 computer with 4 cores (8 threads) and 16 GB of RAM without the use of GPU computing.

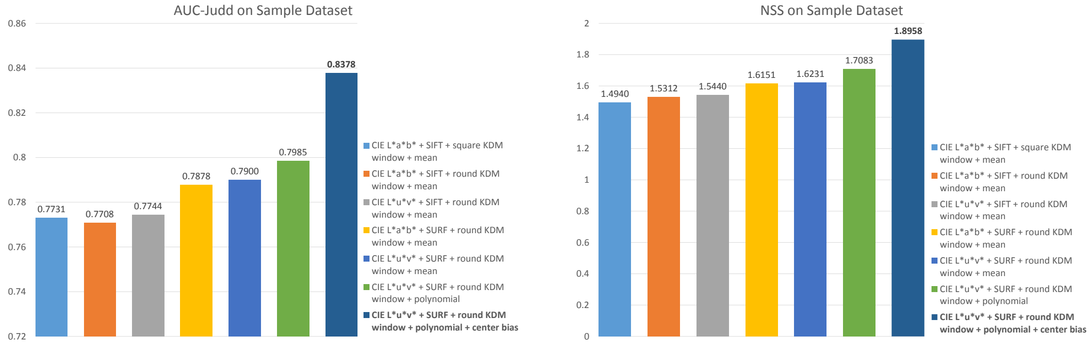
As reported in Tables 3.1, 3.2, 3.3, 3.4, 3.5, our method always got excellent NSS results when compared to traditional unsupervised methods on MIT1003, CAT2000 and EyeCrowd dataset, and exceeded Ensembles of Deep Networks' NSS on the same datasets; comparable results as the other saliency methods have been achieved on ETTO dataset.

We also point out that the new changes we proposed in this work allowed us to outperform our previous results. The CIE L*u*v* + SURF + center bias algorithm performed equally or better than other variations on the sample dataset because it manages to detect salient features with a reduced number of keypoints (Fig. 3.8).

The performances of "mixed" methods (CIE L*a*b* + SURF and CIE L*u*v* + SIFT) have not been reported in the Tables because they were always lower then those of the above-mentioned method.

Furthermore, we want to remark that, despite the different composition of the image datasets we used for our experiments, our method showed quite consistent performances over them and with respect to the various comparison methods.

On average, the execution time required to calculate the saliency map of a 1920x1080 RGB image is 15.4071 s (the algorithm runs as a single thread without any GPU assistance, and its speed could be increased taking advantage of faster mathematical libraries, multiprocessing and GPU computation).



(a) (b)

Figure 3.8: Performance graphs of some variations of our saliency model on the sample dataset.

Table 3.1: Results in various metrics on ETTO dataset.

Saliency method	AUC-Borji	AUC-Judd	CC	KL[1]	NSS	SIM
Our new method	0.9363	0.9518	0.6844	0.8904	4.4399	0.4993
Our previous method	0.8253	0.9208	0.6982	2.0256	4.9197	0.5987
Itti-Koch-Niebur	0.9296	0.9257	0.4670	2.0465	3.1424	0.2050
GBVS	0.9344	0.9511	0.7158	0.8468	4.7342	0.5062
eDN	**0.9494**	**0.9556**	0.4171	2.3551	2.8225	0.1480
SAM-VGG	0.9089	0.9520	0.7061	0.7705	4.8767	0.5685
SAM-ResNet	0.9164	0.9534	**0.7814**	**0.6412**	**4.9985**	**0.6134**
Gaussian center	0.8807	0.9356	0.2080	3.2756	1.3563	0.0726

[1] Lower is better.

Table 3.2: Results in various metrics on MIT1003 dataset.

Saliency method	AUC-Borji	AUC-Judd	CC	KL[1]	NSS	SIM
Our new method	0.7953	0.8151	0.4169	1.3739	1.4071	0.3519
Our previous method	0.7304	0.7459	0.2971	1.5973	1.0065	0.3206
Itti-Koch-Niebur	0.7623	0.7750	0.3307	1.4805	1.1029	0.3226
eDN	**0.8471**	0.8579	0.4096	1.5453	1.2969	0.2976
GBVS	0.8151	0.8288	0.4175	1.2969	1.3819	0.3627
SAM-VGG	0.8409	0.9197	0.8322	0.7745	3.1733	0.6733
SAM-ResNet	0.8442	**0.9267**	**0.8689**	**0.7363**	**3.3384**	**0.7126**
Gaussian center	0.7988	0.8162	0.3288	1.6990	1.0089	0.2691

[1] Lower is better.

Table 3.3: Results in various metrics on CAT2000 dataset.

Saliency method	AUC-Borji	AUC-Judd	CC	KL[1]	NSS	SIM
Our new method	0.7800	0.8005	0.4946	1.0046	1.2644	0.5073
Our previous method	0.7536	0.7660	0.4157	1.1018	1.0658	0.4850
Itti-Koch-Niebur	0.7578	0.7667	0.4131	0.9715	1.0625	0.4647
GBVS	0.7900	0.8012	0.4864	**0.8615**	1.2458	0.4982
eDN	**0.8426**	**0.8510**	0.4933	0.9990	1.2145	0.4477
SAM-VGG	0.7516	0.8394	0.6280	1.1121	1.7642	0.5764
SAM-ResNet	0.7619	0.8438	**0.6550**	1.1586	**1.8256**	**0.5910**
Gaussian center	0.8134	0.8363	0.4504	1.1785	1.0805	0.4021

[1] Lower is better.

Table 3.4: Results in various metrics on EyeCrowd dataset.

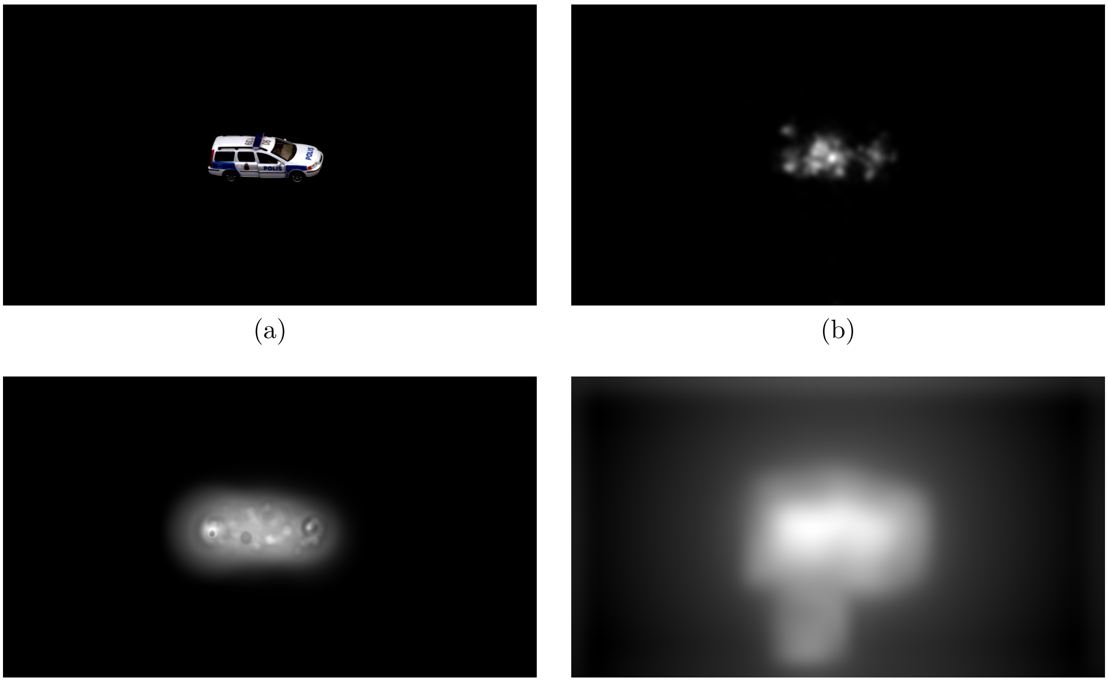
Saliency method	AUC-Borji	AUC-Judd	CC	KL[1]	NSS	SIM
Our new method	0.7737	0.7842	0.4959	**0.7852**	1.1020	0.5144
Our previous method	0.7251	0.7333	0.3745	0.9294	0.8324	0.4742
Itti-Koch-Niebur	0.6544	0.6993	0.2786	1.0463	0.5843	0.4442
GBVS	0.7562	0.7641	0.4444	0.8420	0.9671	0.4940
eDN	**0.7832**	0.7902	0.4709	0.9517	1.0068	0.4552
SAM-VGG	0.7409	0.8429	0.7064	1.5619	1.8898	0.6397
SAM-ResNet	0.7495	**0.8466**	**0.7242**	1.4527	**1.9682**	**0.6533**
Gaussian center	0.7285	0.7373	0.3770	1.0631	0.7788	0.4242

[1] Lower is better.

Table 3.5: Results in various metrics on FIGRIM dataset.

Saliency method	AUC-Borji	AUC-Judd	CC	KL[1]	NSS	SIM
Our new method	0.7794	0.8022	0.4315	1.1692	1.2567	0.4065
Our previous method	0.7094	0.7270	0.2932	1.5519	0.8601	0.3718
Itti-Koch-Niebur	0.6700	0.7470	0.2820	1.3278	0.8220	0.3717
GBVS	0.8076	0.8197	0.4470	**1.0519**	1.2599	0.4300
eDN	**0.8621**	**0.8721**	0.5031	1.2348	1.3795	0.3735
SAM-VGG	0.7739	0.8699	0.6166	1.1518	2.0299	0.5373
SAM-ResNet	0.7835	0.8696	**0.6242**	1.1524	**2.0538**	**0.5397**
Gaussian center	0.8297	0.8577	0.4291	1.3717	1.1409	0.3435

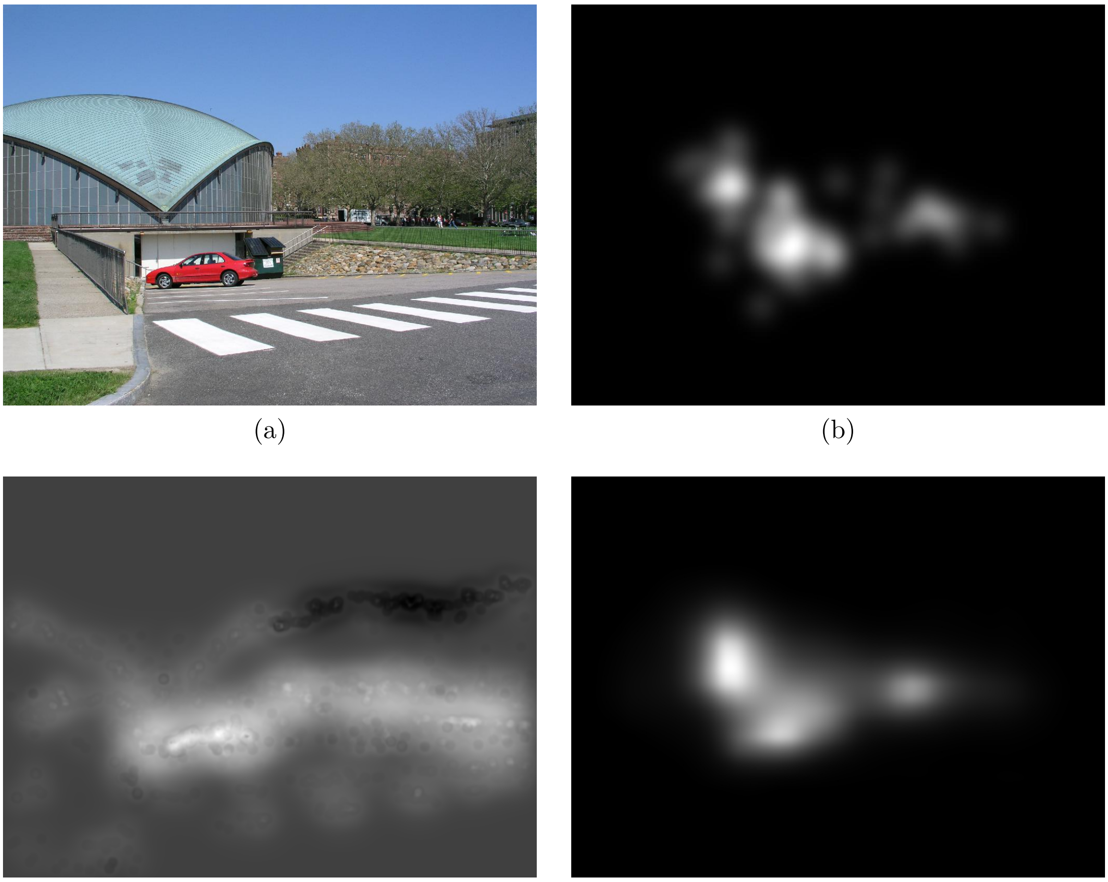
[1] Lower is better.



(a) (b)

(c) (d)

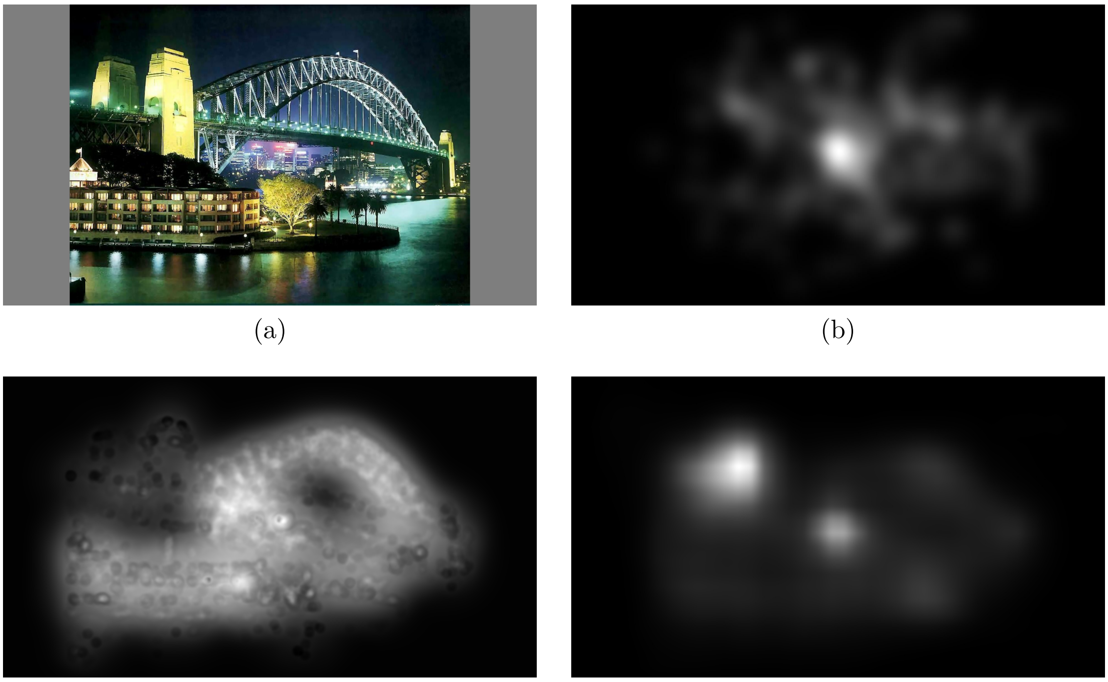
Figure 3.9: (a) An image from the ETTO dataset, (b) its fixation map, (c) its saliency map generated with our multi-scale color SURF KDM method, and (d) its saliency map generated with the eDN method.



(a) (b)

(c) (d)

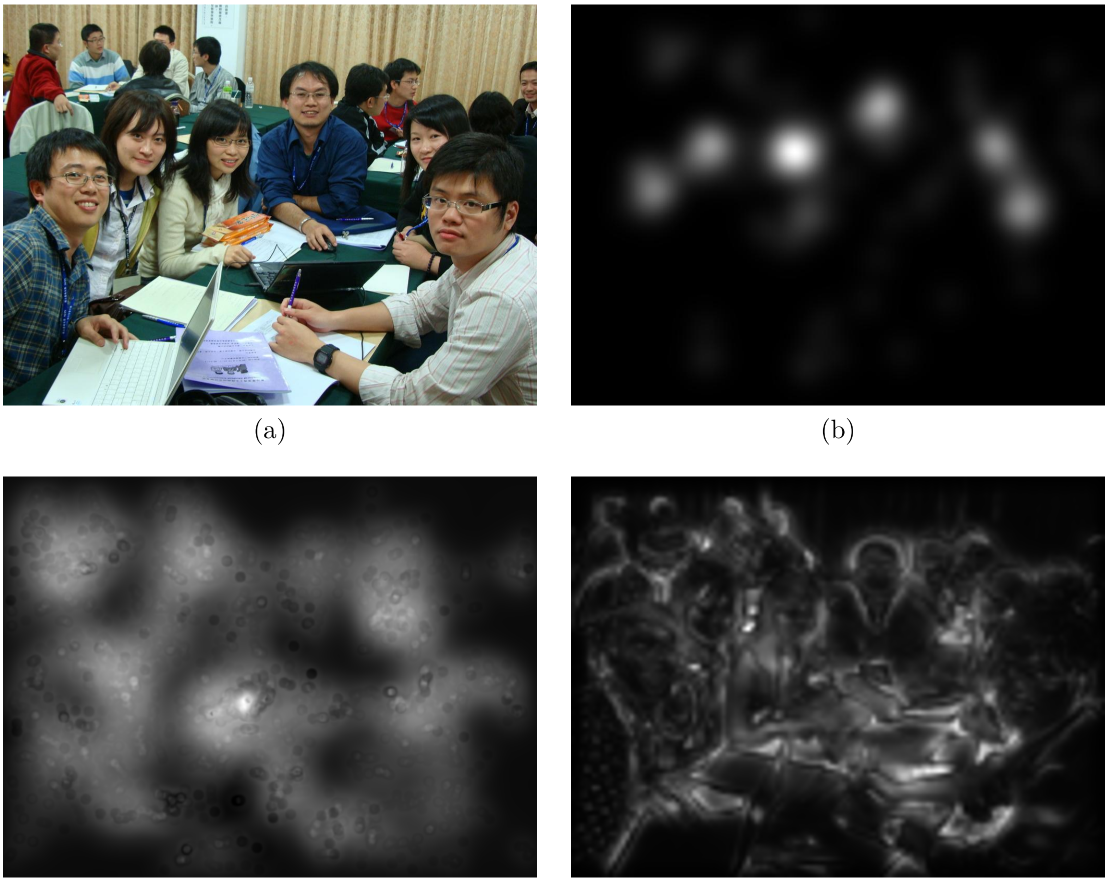
Figure 3.10: (a) An image from the MIT1003 dataset, (b) its fixation map, (c) its saliency map generated with our multi-scale color SURF KDM method, and (d) its saliency map generated with the SAM-ResNet method.



(a) (b)

(c) (d)

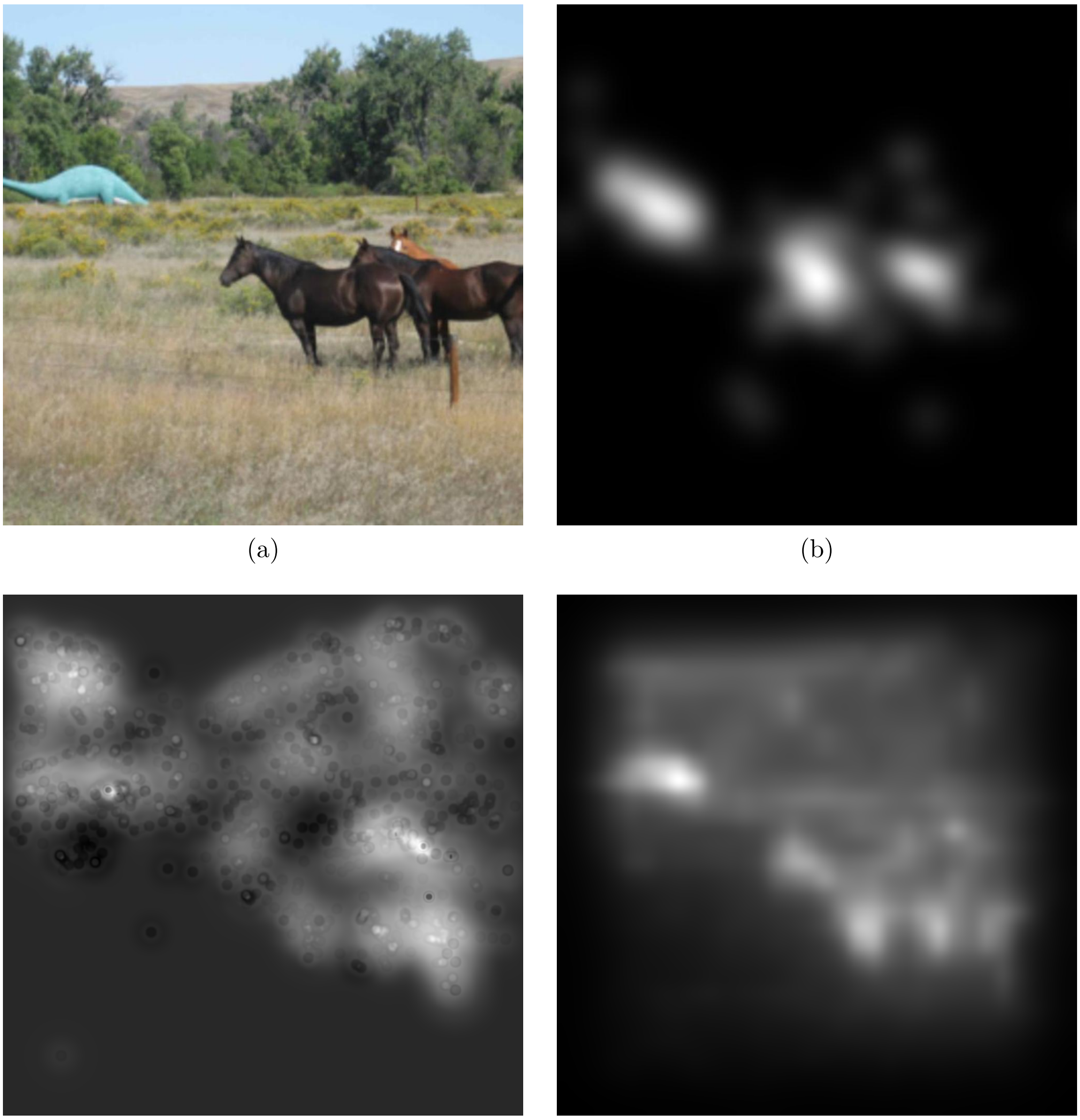
Figure 3.11: (a) An image from the CAT2000 dataset, (b) its fixation map, (c) its saliency map generated with our multi-scale color SURF KDM method, and (d) its saliency map generated with the SAM-VGG method.



(a) (b)

(c) (d)

Figure 3.12: (a) An image from the EyeCrowd dataset, (b) its fixation map, (c) its saliency map generated with our multi-scale color SURF KDM method, and (d) its saliency map generated with the Itti-Koch-Niebur method.



(a) (b)

(c) (d)

Figure 3.13: (a) An image from the FIGRIM dataset, (b) its fixation map, (c) its saliency map generated with our multi-scale color SURF KDM method, and (d) its saliency map generated with the GBVS method.

3.5 Image Content Enhancement through Salient Regions Segmentation for People with Color Vision Deficiencies

In this section we describe the steps of our work, starting from the eye-tracking session we aim at determining the most peculiar differences between normal and color deficient vision systems with respect to a fixed number of images, then we tackle the segmentation and the re-coloring of the regions with different saliency levels; at last a further eye-tracking session assesses the enhancement of the images. We point out that we are interested in detecting differences in human visual system behavior among people with normal vision system and people affected by color vision deficiencies (in Fig. 3.14 only dichromatic people will be able to easily recognize the word NO standing out from the background).

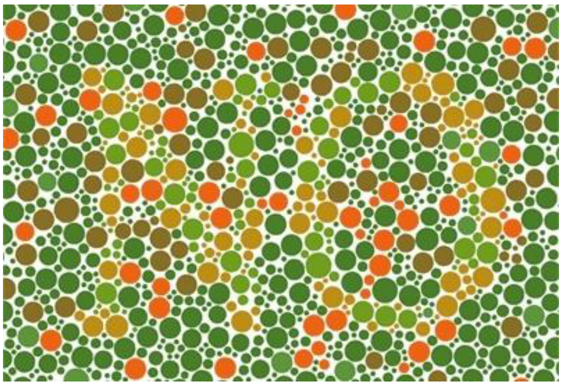


Figure 3.14: Unlike people with normal vision system, people with dichromatic vision system are able to easily recognize the word "NO".

3.5.1 Eye-tracking Session

The experimental sessions involved 8 subjects with normal vision system and 8 subject with color deficient vision system. More in detail, 3 subjects were affected by deuteranopia and 5 subjects were affected by protanopia. We conducted two experimental eye-tracking sessions, the first is focused on detecting how different are the fixation points among color blind and normal people, the second is needed to assess the effectiveness of our method in enhancing the images for color blind people.

Figure 3.15: Images taken during the eye-tracking session: starting from the calibration (the far-left image) the eye-tracker records the eye movements, the saccadic movements and the scanpaths.

Both eye-tracking sessions consist of repeating the same procedures, but the first session also includes a test with Ishihara plates in advance to evaluate which kind of color vision deficiency the subjects (EnChroma Inc., 2010) are affected by. The experimental sessions have been conducted in a half-light room, the subjects were kept at a distance of almost 70 cm from a 22 inches monitor with a spatial resolution of 1920 x 1080 pixel (Fig. 3.15). During the eye-tracking session a Tobii EyeX device recorded the eye movements, the saccadic movements and the scanpaths of each subject while he or she was looking at the images shown on the screen. For each subject a calibration step was needed to minimize saccadic movement tracking errors. Each image was shown on the screen for a time of 3 s during which the Tobii EyeX acquired spatial coordinates of the eye movements (at a mean rate of 160 spatial coordinates per 3 s because of the sampling of about 55 Hz). Before switching to the next image the screen was turned gray for one second to refresh the observer retina from the previous image signal. The eye-tracking session procedures follow the same overall scheme of previous scientific works (Ardizzone et al., 2011, 2013a, 2017a) focused on visual saliency studies. Each session lasted approximately for 7 minutes per subject. For the purpose of our experiments, we created an ad-hoc dataset by merging almost 90 images from different public datasets (MIT1003 (Judd et al., 2009a), CAT2000 (Borji and Itti, 2015); NUSEF (Ramanathan et al., 2010); MIT300 (Bylinskii et al., 2015b)): it consists of images containing meals, plants, objects, fruits, people, portraits, animals, pets, synthetic pictures showing texture patterns. All the images have in common a prevalence of red-green chromatic contrasts (of interest for protanopia and deuteranopia deficiencies); we did not take into account images with yellow-blue chromatic contrast because we did not have available people affected by tritanopia vision deficiency. All the fixation point maps we collected during two experimental sessions have been gathered into a public available ground-truth under the name of Eye-Tracking of Colour Vision Deficiencies (EToCVD) (Bruno, A. and Gugliuzza, F. and Ardizzone, E. and Giunta, C. and Pirrone, R., 2018).

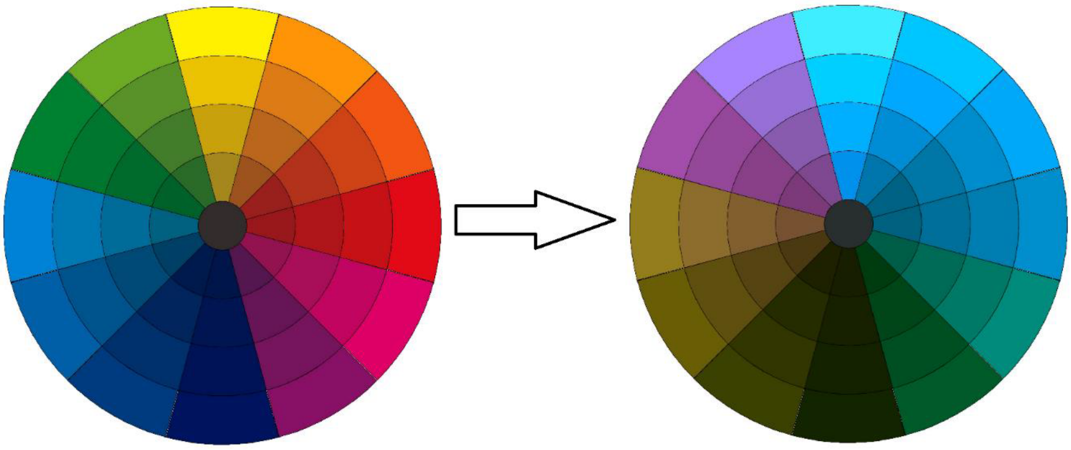


Figure 3.16: RGB to CIE L*a*b* conversion allows us to manage with color mapping within color frequencies well perceived by color blind people.

The eye movement data reveal the locations of the images looked at by the observer during the experimental session. The fixation points are computed by averaging the spatial x and y coordinates of each eye movement (left and right eye movements). The coordinates are converted to the range of the spatial resolution of the screen. Each time a pixel is observed its value is incremented starting from zero. Then the saliency map is smoothed through a Gaussian convolution. In Fig. 3.17 it is shown how different the fixation points are between observers affected by protanopia and observers with normal color vision system. It is remarkable how people affected by protanopia miss several details because they fall within the color spectrum they are not able to discriminate. We want to point out that we aim to enhance the content of the image to make both people affected by protanopia and deuteranopia able to detect details they cannot detect because of the color blindness constraint.

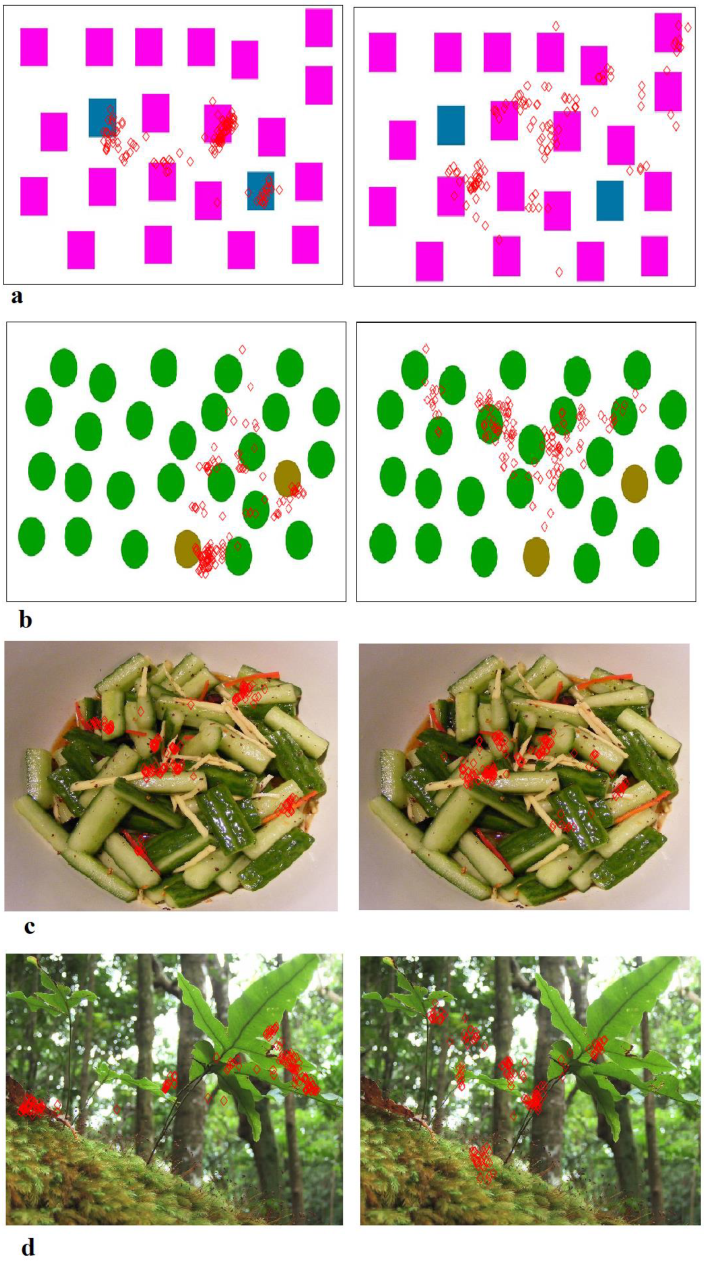


Figure 3.17: The visual perception of an image can be represented by the fixation points (red diamonds overlaid on the images) for both normal vision system (left column) and color blind vision system (right column). Some details are missed by people with color vision deficiencies and this is revealed by the lack of fixation points on the details noticed by people with normal vision system.

3.5.2 Proposed Method

For our purpose we put our effort on assessing the usage of visual saliency in content enhancement for color blind people.

In our approach, for instance, the saliency map is extracted from an image by using the algorithm we proposed in our previous work (Ardizzone et al., 2017a) based on the spatial distribution of local keypoints. Since we wanted to investigate the relation between the saliency and the color information, we tackled the saliency detection on CIE L*a*b* color space because of the independence of the luminance channel from color channels (a* and b*). A saliency map is computed along each channel of the CIE L*a*b*, then the output saliency map is obtained by averaging the saliency maps of each channel.

A color vision deficiency simulation method (Milić et al., 2015), in the same way of Vienot et al. (Viénot et al., 1999), is applied to have a dichromatic version of the original image. Vienot's method allows us to choose the color deficiency to be simulated through a function parameter. Afterward, the saliency map of the simulated dichromatic version of the image is extracted by using the same procedure as above (Ardizzone et al., 2017a). The saliency error is computed as the absolute differences between the two saliency maps. We consider the saliency error as a computational measure of how color vision deficient system behavior is different from normal vision system behavior. The saliency error is used as a weight and multiplied with the difference between the original image and the simulated dichromatic version, then the result is converted to the RGB space. A correction vector is multiplied with the resulting RGB image and an average function is applied along the three RGB channels giving rise to a single map, furthermore a 3x3 sized Gaussian filter is applied to smooth the map noise.

The smoothed map is segmented by using the adaptive Otsu segmentation (Otsu, 1979). The saliency error is then represented as segmented regions. For the enhancement purpose only the segmented regions have been taken into account, that is, the pixels of the segmented regions are transformed with a negative mapping as in the equations reported in (3.12), that represent a rotation of 180° of the a* and b* channels in CIE L*a*b* space:

$$\begin{aligned} L^{*'} &= L^{*} \\ a^{*'} &= -a^{*} \\ b^{*'} &= -b^{*} \end{aligned} \tag{3.12}$$

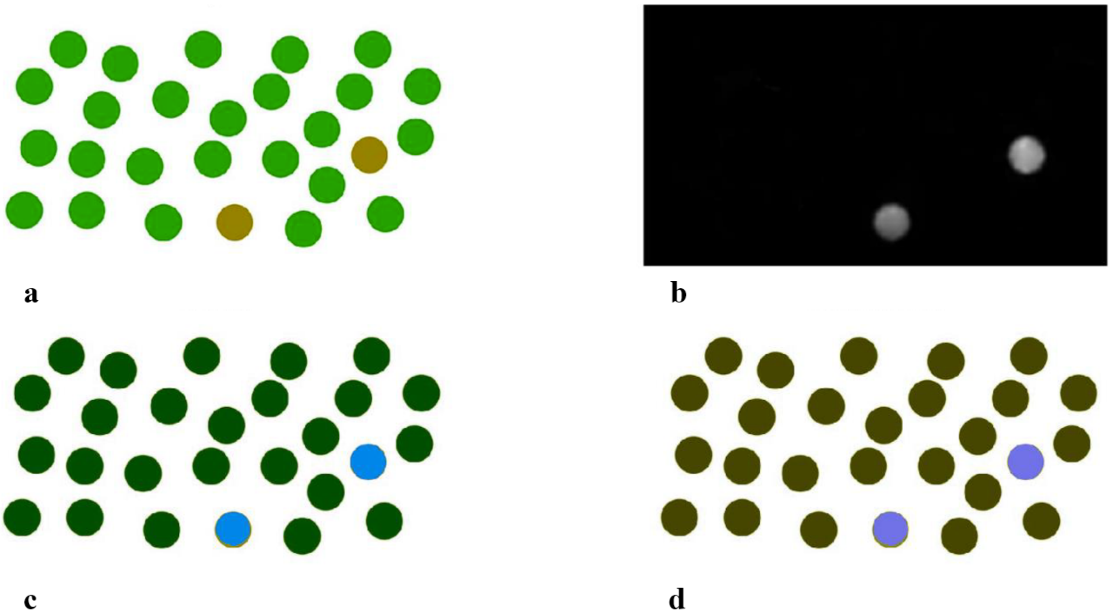


Figure 3.18: (b) The saliency error is computed as the difference of the saliency maps of (a) the original image and the color blind version of the image. (c) The saliency error regions are segmented and color boosted on CIE L*a*b* color space by using the opposite value with the a* and b* channels, (d) the enhancement is also mapped in the color blind domain.

Using the above equations and then converting the resulting image back to RGB color space we noticed that at first sight color blind people were able to perceive more details from the regions with pixels falling within the color frequency that they were not able to discriminate before. More in detail, as you can see from Fig. 3.16 and the equations reported above, hues close to red are mapped to hues very close to blue spectrum. Before the aforementioned processing steps people affected by protanopia and deuteranopia were not able to discriminate red edges over green background. A new eye-tracking session has been conducted to assess the effectiveness of the aforementioned enhancement and processing steps with the visual perception of color blind people.

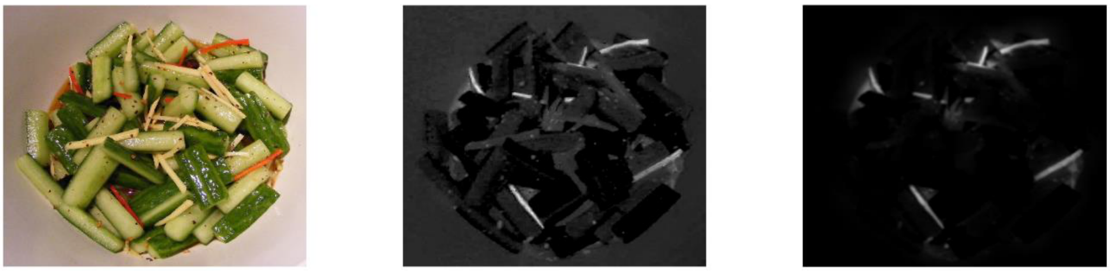


Figure 3.19: Perceptual differences highlighting. For a given image (left) some enhancement methods use the average of the differences of L*a*b* channels between the original image and the version with simulated color vision deficiency (center). We adopted the difference of L*a*b* channels between the original image and the version with simulated color vision deficiency, weighted by the saliency difference (right).

3.5.3 Experimental Results

In this section we want to describe our findings with respect to the behavior of people affected by protanopia and deuteranopia, after the we performed the enhancement processing steps highlighted in the previous section. Once the images had been enhanced, after 20 days we repeated the eye-tracking session with subjects affected by color vision deficiencies. The idea behind a second experimental session is to assess the effectiveness of our enhancement approach by comparing the real fixation points acquired before and after the enhancement step. We considered the real fixation point maps of the observers with normal color vision system as our ground truth. First we analyzed the differences between the real fixation point maps of color blind people and our ground truth, both acquired during the first eye-tracking experimental session, and we used that as reference for our comparison studies. Then we analyzed the differences between real fixation point maps acquired during the second eye-tracking session and the ground truth. We conducted experiments with people affected by protanopia and deuteranopia and we collected the real fixation point maps to be evaluated with metrics such as NSS and AUC focused on visual perception processes.

Very interesting results can be observed from the images in Fig. 3.20. The observers affected by protanopia were able to discriminate and notice some details they did not look at before (Fig. 3.16). In spite of some imperfections in the re-coloration of the segmented region (Fig. 3.20a) we noticed that the overall distribution of the fixation points (Fig. 3.20b, 3.20c, 3.20d) acquired during the second eye-tracking session is closer to the corresponding ground truth map than the distribution of the fixation points acquired during the first session.

As shown in Fig. 3.21a, 3.21b, 3.21c, 3.21d the eye movements of the observers affected by protanopia and deuteranopia can be really different to those of people with normal color vision system. The interesting thing is that, analyzing the improvements obtained with our enhancement method by observing the fixation points map of subjects affected by protanopia (Fig. 3.22c) and deuteranopia (Fig. 3.22d), the improvement is noticeable because the fixation points are quite closer to our ground truth (Fig. 3.21a).

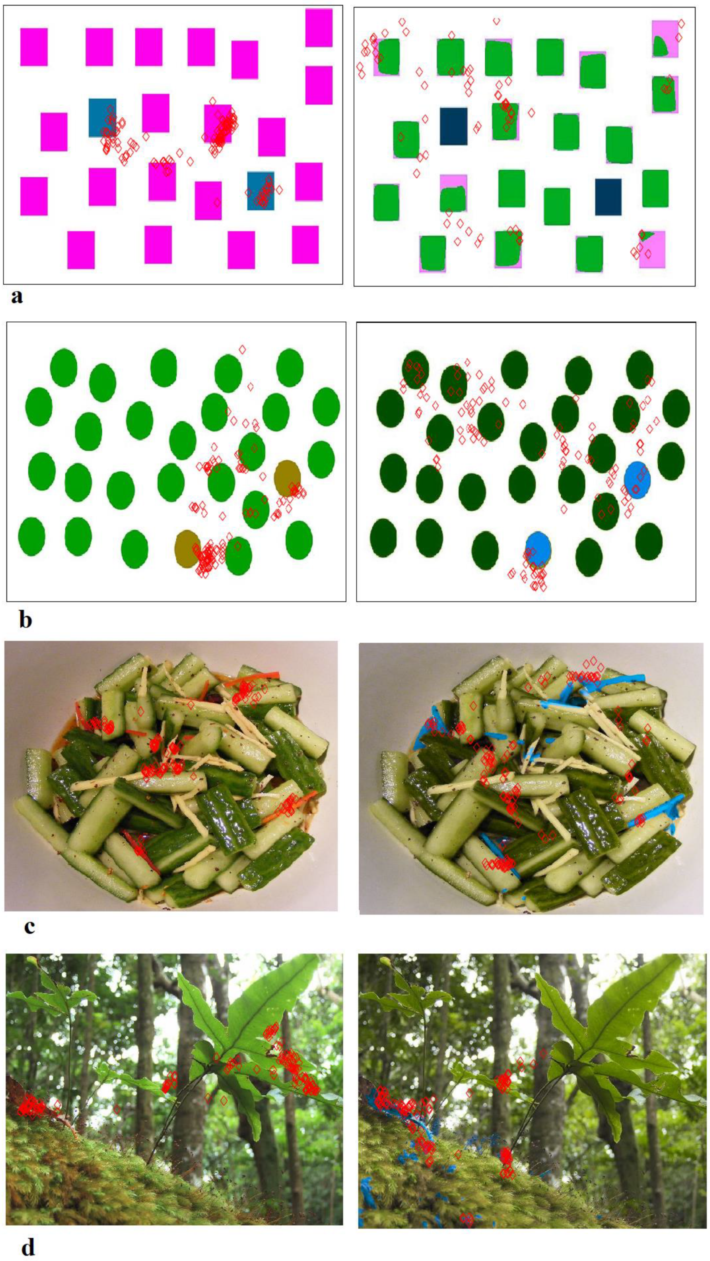


Figure 3.20: The fixation points (red diamonds overlaid on the images) of observers with normal color vision system (left column) and the fixation points of people affected by protanopia. The images from the right column are from the second eye-tracking session when observing the images enhanced by our method.

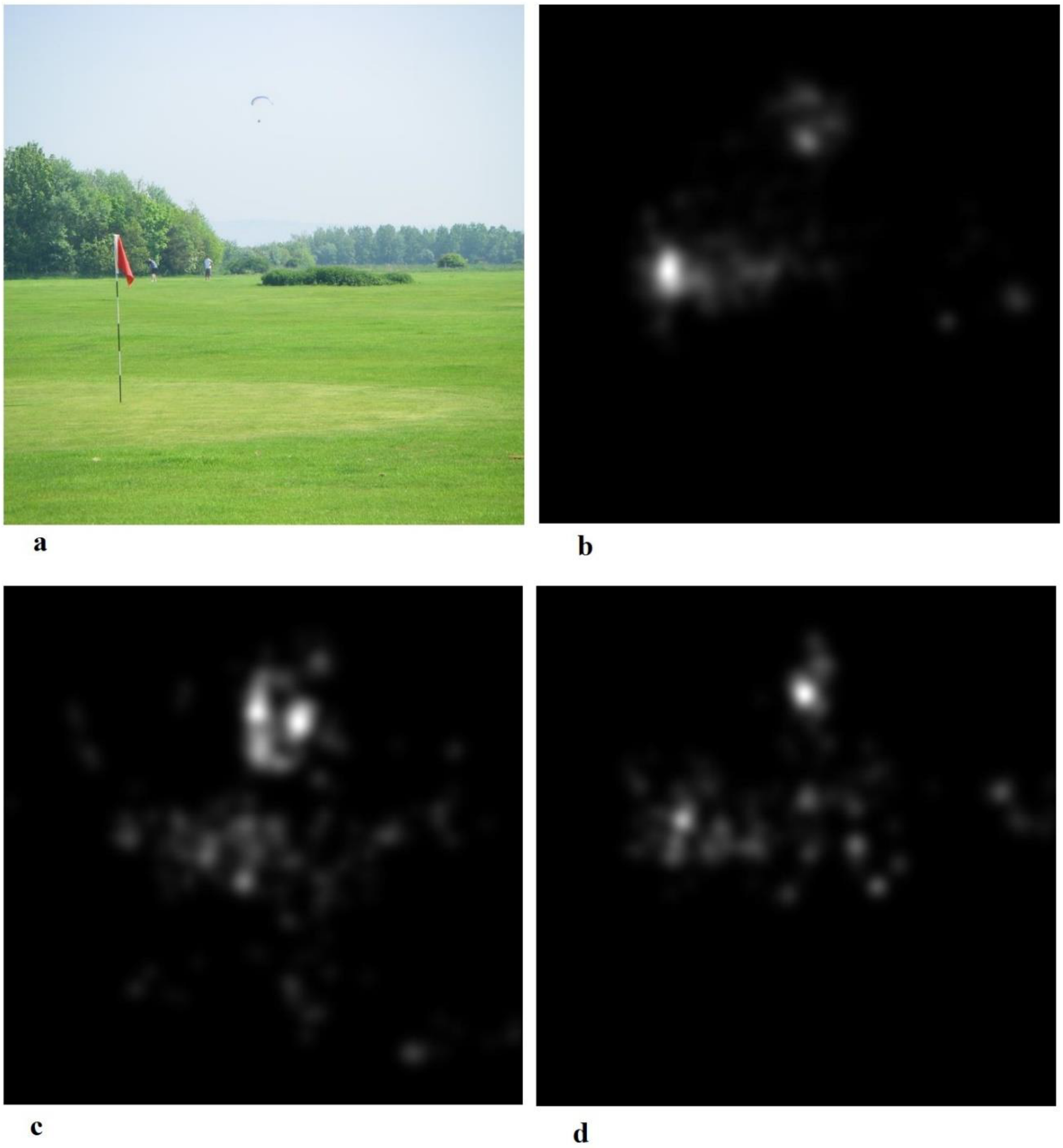


Figure 3.21: For (a) a given image we collected the fixation points from (b) a normal observer, (c) an observer affected by protanopia and (d) an observer affected by deuteranopia.

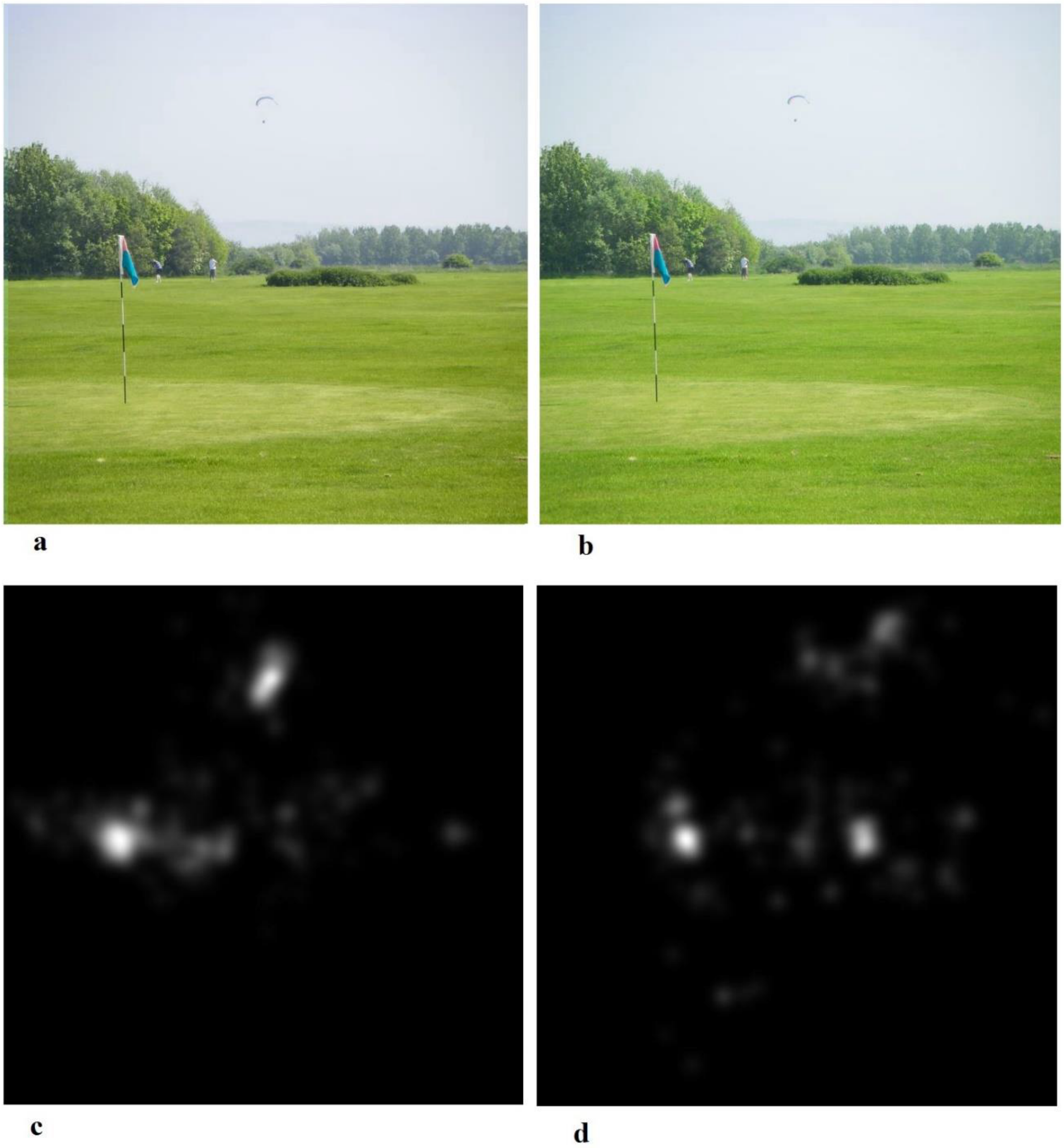


Figure 3.22: The enhancement assessment on (a, b) the images is supported by the fixation point maps for observers with (c) protanopia and (d) deuteranopia.

The performance of our method is shown by using metrics such as AUC and NSS well suited for quantifying how close are the fixation point maps of color blind people to the fixation point maps of people with normal vision system. For the sake of clarity, we studied the performance of our method on people affected by protanopia and deuteranopia separately.

We want to point out that scientific literature on visual attention revealed that during the first 200 ms of image observation humans tend to fix locations around the image center, and this results in a center biased fixation point map. It is of our interest to analyze the experimental results by distinguishing two case studies:

- Collecting fixation point data of the observer over the entire time interval for each image (3 s)
- Collecting fixation point data from 200 ms to 3 s.

The objective of excluding the first 200 ms from the fixation point data is to have unbiased data to be analyzed.

Normalized Scanpath Saliency (NSS) allows us to give a measure of how close a saliency map is to a real fixation point map. The metric was originally devised to compute the distance between a computational saliency map and a real fixation point map. In our method we used NSS to compute the distance between the real fixation point map of people with normal vision system and the real fixation point map of people with color vision deficiencies. NSS metric gives us a scalar value. A NSS value of zero means the maps are very different, conversely, a higher NSS value means higher similarity between the maps.

Area Under Curve metric (AUC) is computed as the area of the ROC curve (Receiver Operator Characteristic). It is a scalar representation of the predicted performances of a classifier. AUC value falls within the range [0, 1].

Looking at the AUC and NSS results of the protanopia case study (Fig. 3.23, 3.24) we noticed that the enhancement of the images allowed observers to detect more details previously falling in the color blind spectrum. Excluding the data acquired during the first 200 ms for the reasons explained above, our results reached an average score increase of approximately 0.08 AUC and 0.5 NSS. In Fig. 3.27 we show some examples with different AUC and NSS values related to quite meaningful images with the corresponding eye-tracking fixation point map of the first eye-tracking session and the fixation point map related to the second eye-tracking session, giving us a visual and qualitative demonstration of the improvement we achieved.

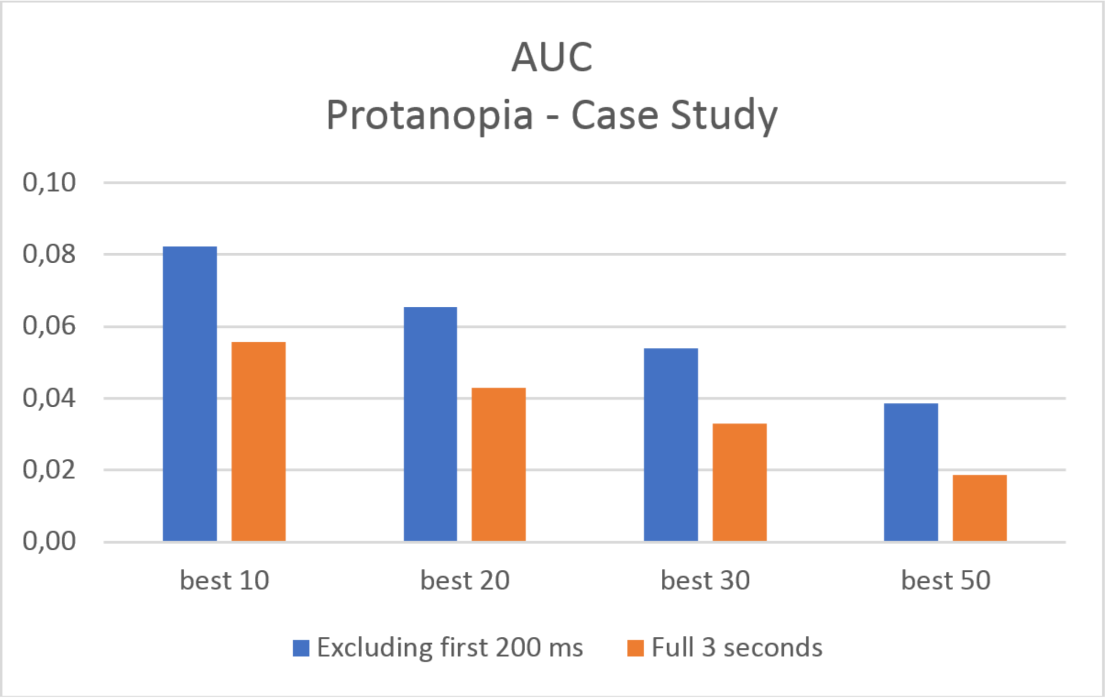


Figure 3.23: Average AUC score increase of the best 10, 20, 30, 50 cases. Protanopia case study.

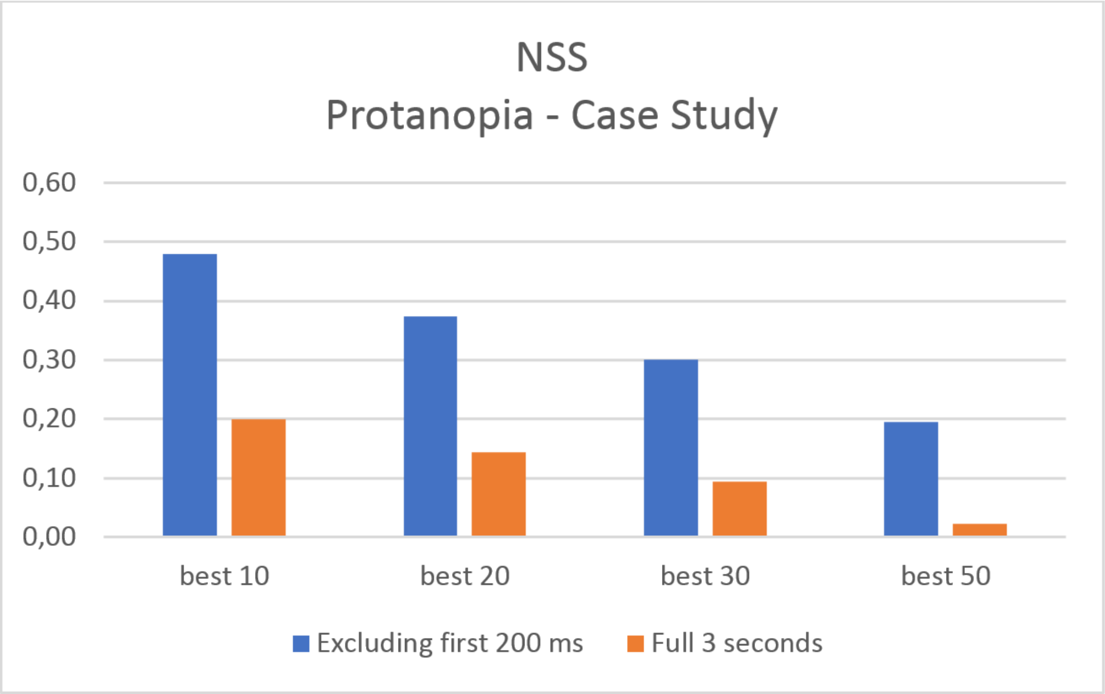


Figure 3.24: Average NSS score increase of the best 10, 20, 30, 50 cases. Protanopia case study.

In Fig. 3.25, 3.26, 3.28 we plot the histogram graph of AUC and NSS average score increase with respect to the deuteranopia case study and we show meaningful images and the corresponding fixation point maps. It is noticeable that in the case of observers affected by protanopia we reached an average score increase of approximately 0.05 AUC and 0.3 NSS (both excluding the first 200 ms).

As can be deduced from the histogram bars, there are a lot of differences between including and excluding the first 200 ms in our case study. It is evident that observers affected by protanopia show a more peculiar center bias in their visual attention path.

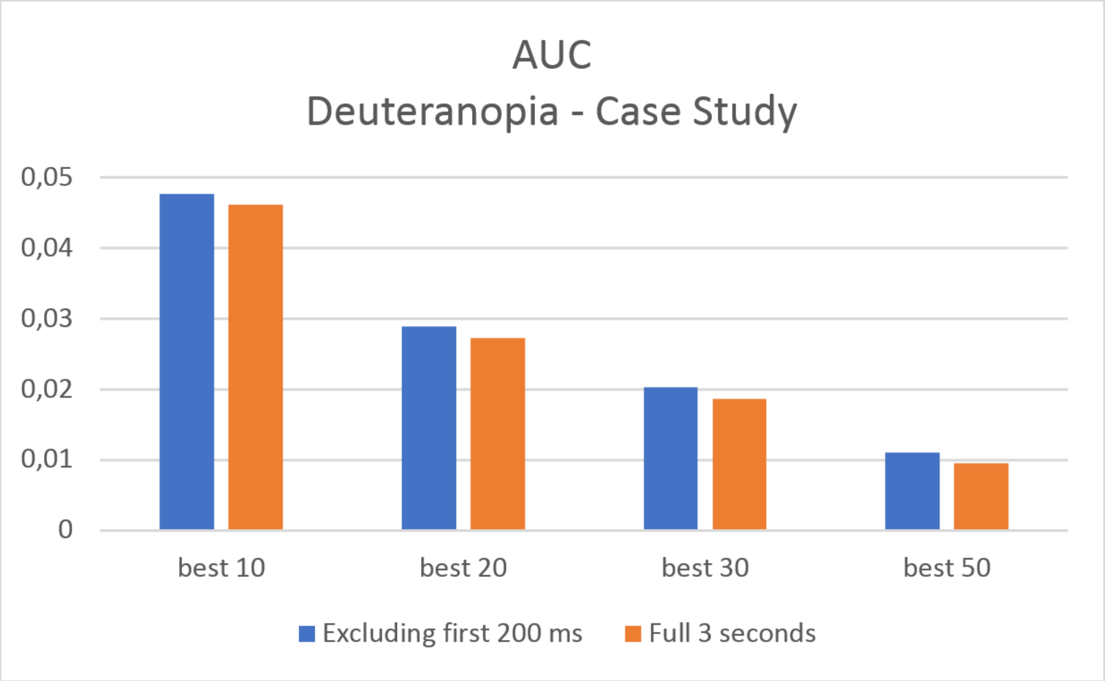


Figure 3.25: Average AUC score increase of the best 10, 20, 30, 50 cases. Deuteranopia case study.

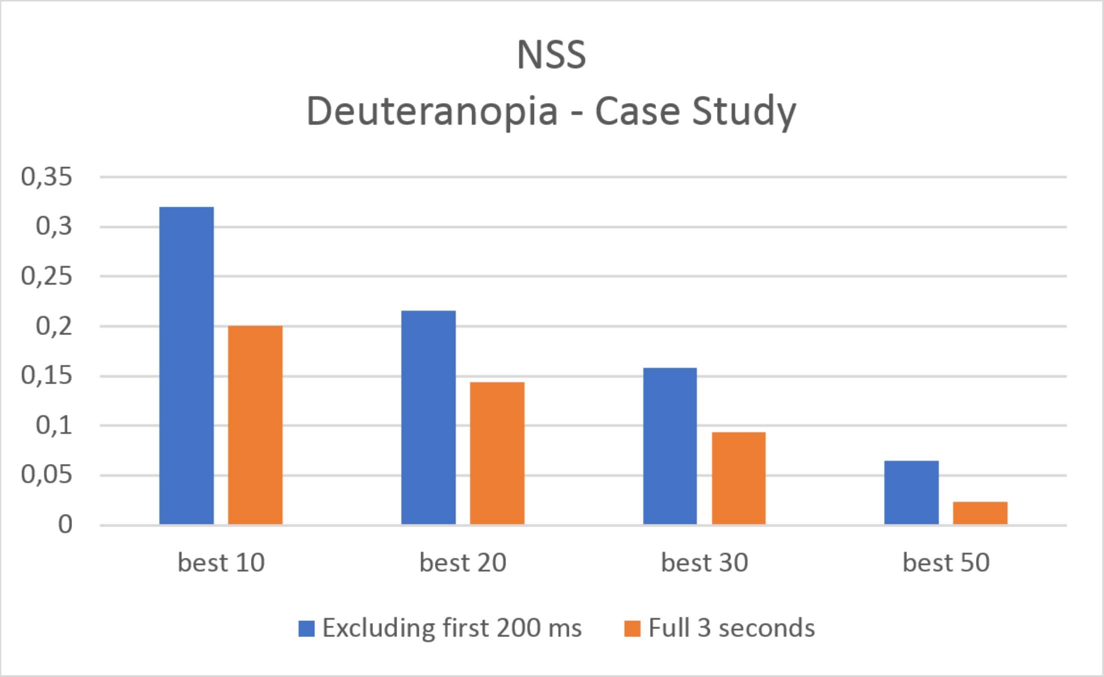


Figure 3.26: Average NSS score increase of the best 10, 20, 30, 50 cases. Deuteranopia case study.

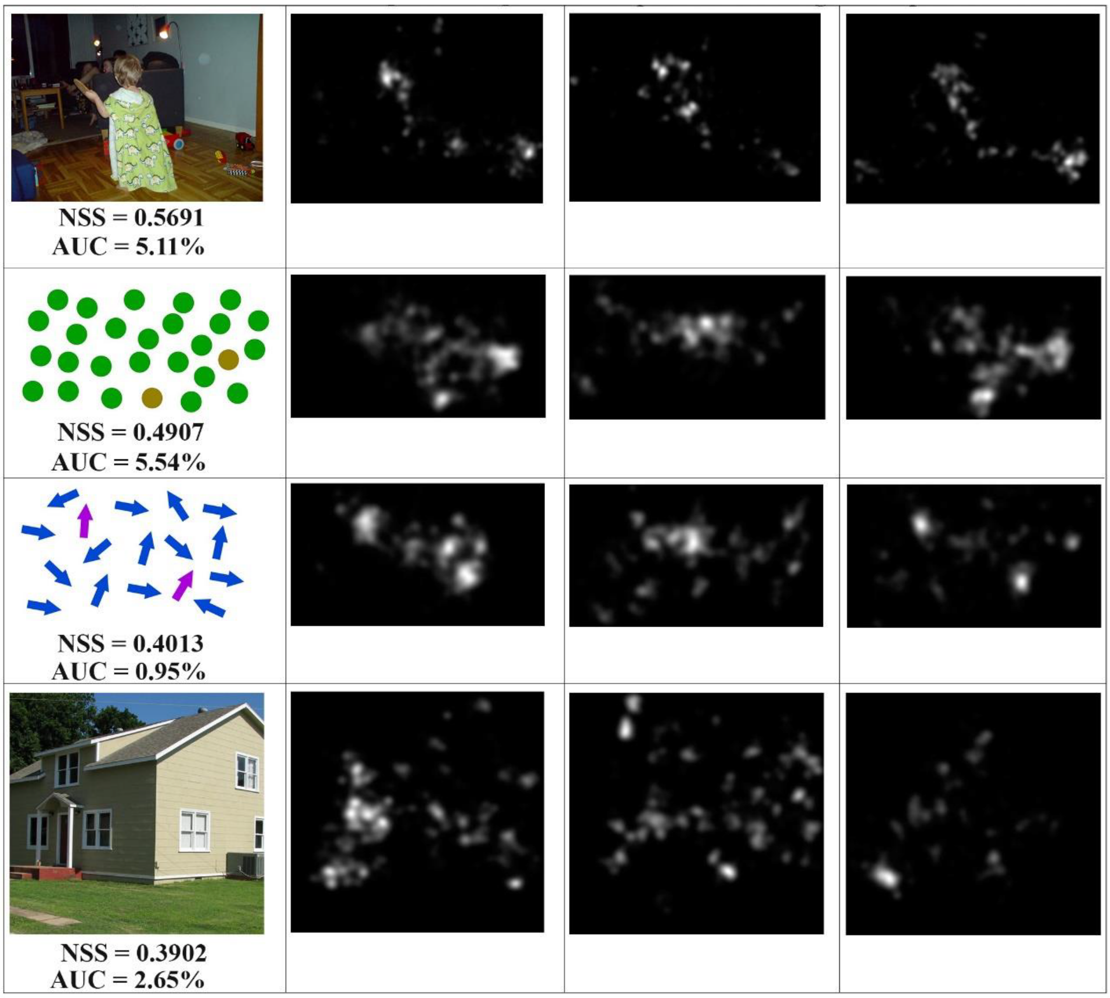


Figure 3.27: For a given image (first column) we collected the fixation points from normal observer (second column), from observers affected by protanopia (third column), from observers affected by protanopia looking at the enhanced image (fourth column).

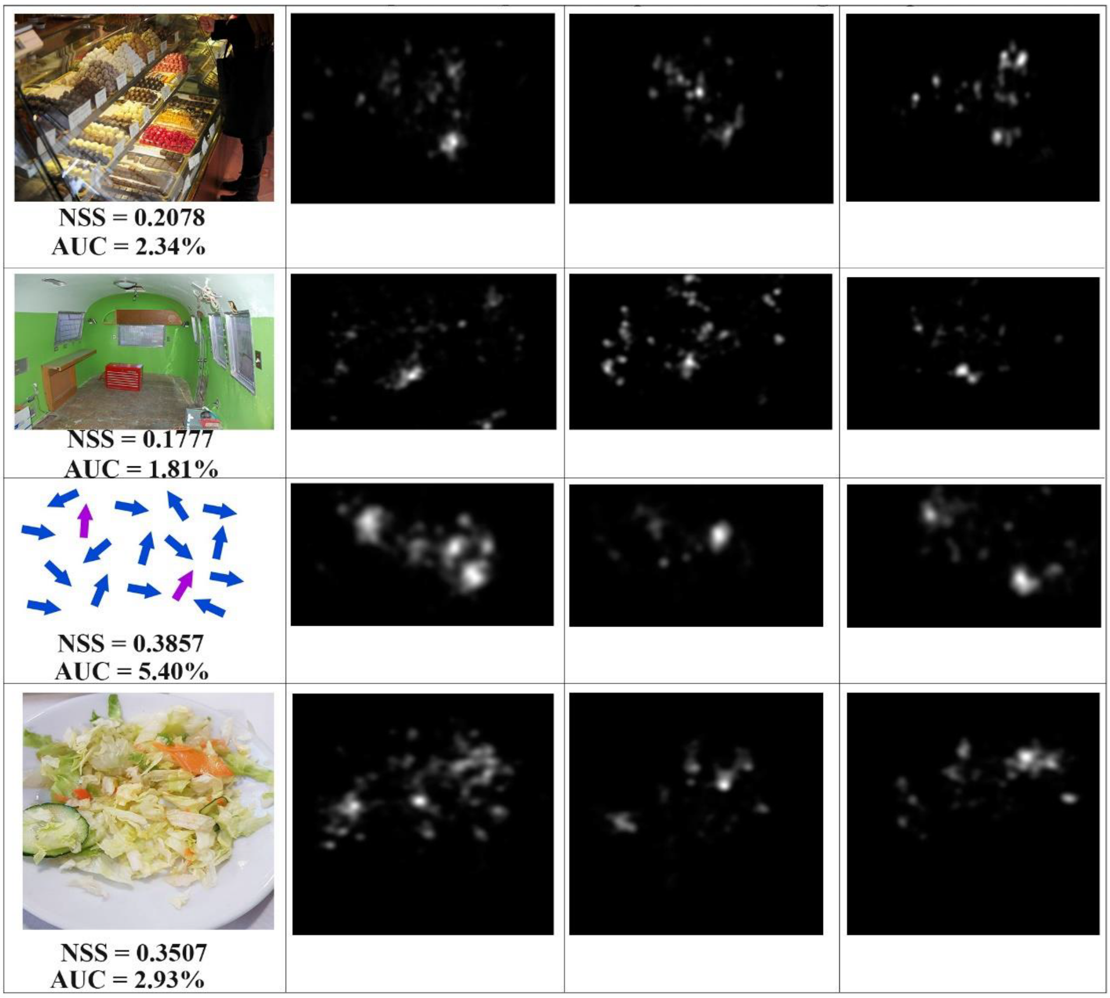


Figure 3.28: For a given image (first column) we collected the fixation points from normal observers (second column), from observers affected by deuteranopia (third column), from observers affected by deuteranopia looking at the enhanced image (fourth column).

The experiments have been conducted by using a Tobii EyeX eye-tracker recording the eye movements at a sampling rate of about 55 Hz; the data have been processed in MathWorks MATLAB: in greater detail we dealt with Tobii EyeX calibration and eye-tracking parameter tuning inside TobiiMatlabToolbox 3.1 (Gibaldi et al., 2017, 2016). We used a workstation with a quad-core 2.4 GHz processor and 16 GB of RAM for our experiments.

Chapter 4

Conclusions

4.1 Satellite Remote Sensing

In this work we proposed a complete system for weather satellite data reception. It is able to predict and track satellite passes, receive and demodulate both legacy NOAA APT signals and high resolution data contained in newer signal formats transmitted by various satellites. Despite the new system being still a work in progress, its specifications already match or exceed those of commercially available solutions (Dartcom Systems Ltd., 2018), while being very easily reconfigurable to support new signal and data formats. In the future we want to improve our solution with modular upgrades, thus adding the following:

- Multiple feeds and LNAs for multiple frequency bands
- Support for various meteorological signal and payload formats
- Support for error correction codes (where available)
- An algorithm for thermal maps generation based on the analysis of multi-spectral infrared images
- Noise profiling based on long time analysis of the radio frequency interferences when the interferences are located near the receiver.

We aim to develop a modular solution for the prediction of atmospheric phenomena, based on the analysis of the images retrieved from several satellites (each one transmitting at its own frequency and with its own modulation scheme) at different times of the day. As far as it concerns cloud detection we proposed a solution that is mainly based on a simple and straightforward clustering method, that is, K-means. The method performs well with $K = 3$ but, generally speaking, the performance of a segmentation method based on clustering depends on the elements inside the image. We think that a a fixed number $K = 3$ for K-means algorithm could not be sufficient for cloud detection whether the image contains elements such as soil, sea, cirrus, aerosols and clouds as well. In future works we want K cluster number to be adaptive with respect to the image we deal with. We could analyze the histogram of a patch to count the number of distribution modes, trying to establish a correspondence with the number of elements inside the image.

All the hardware and software improvements described above will allow us to receive multiband signals from space to enable more complete information to be extracted as infrared, visible and ultraviolet channels.

4.2 Visual Saliency

In this work we reported our findings and research on visual saliency with some methodological novelties with respect to our previous works. We conducted several experiments to evaluate the effectiveness of our method with respect to different state-of the-art methods. The performances have been evaluated over the most popular datasets such as MIT1003, CAT2000, EyeCrowd, FIGRIM, as long as our own dataset, ETTO. We investigated the relationships between saliency and visual attention by comparing the saliency maps and the real fixation point maps, the objective was to show how accurate is the detection of the most important regions from a perceptual viewpoint by using a method based on pattern recognition and image processing techniques. In order to evaluate the accuracy and the effectiveness of our method, we compared the results of our approach on the above mentioned datasets to those of the most popular methods belonging to different classes (deep learning, bottom-up, supervised, unsupervised), getting excellent NSS results against other unsupervised methods and even outperforming a method based on deep learning. The main novelties of this work are the extension of our approach by using multi-scale KDMs on the CIE L*u*v* color space and the comparisons with the other state of the art methods on both generic and specific image categories. Last but not least, we tried to answer the following questions: do saliency methods perform better with respect to general visual attention processes than the specific object attention processes (or vice-versa)? Which method does perform best for eye movement prediction? By comparing the experimental results on ETTO dataset, it has been shown that our approach achieves comparable results as the other saliency methods with respect to the object attention process.

We also believe that it is still worth to investigate cues for visual saliency detection improvements in spite of the excellent accuracy achieved by many state of the art methods. We need to focus on methods who provide a well balanced trade-off between an overall good detection accuracy and acceptable hardware requirements. We have been working on the reimplementation of our novel method in Python, taking advantage of GPU computation frameworks like CUDA (supported by OpenCV). We expect that this code rewrite would greatly speed up the algorithm's execution. This would make it feasible for on-the-fly intra-frame video saliency extraction when used in conjunction with state-of-the-art inter-frame saliency extraction techniques.

We also would like to extend our saliency studies and findings to other visual perception tasks related to focused and sustained attention tasks. For this purpose we will collect different eye-tracking data from viewers performing tasks of the aforementioned types. Furthermore, we have used only traditional supervised learning methods so far, but in the future we will be able to train deep learning systems by using our datasets as training and validation sets. In this case we will focus our research on finding the best network architecture and hyperparameters to achieve the highest accuracy in saliency detection.

4.2.1 Image Content Enhancement through Salient Regions Segmentation for People with Color Vision Deficiencies

The Human Visual System tends to fix some specific points and regions of the image in the first seconds of observation summing up the most important and meaningful parts of the scene. In this work our findings are related to the differences of eye movements with respect to normal and color vision deficient visual systems. Two eye-tracking experimental sessions allowed us to detect and analyze the image details that are not well perceived and fixed by color blind observers. We provided a method to enhance color regions of the image based on CIE L*a*b* color mapping of segmented salient regions. The segmentation is performed by using a saliency weighted difference between the original input image and the corresponding color blind altered image. A second eye-tracking session with color blind people on the enhanced images revealed that the real fixation points are then more coherent with the normal visual system: up to 10% for people with protanopia, up to 5% for people with deuteranopia. The method we proposed makes color blind people able to detect some more red-green details from the images with respect to the original image. We are now working to improve the method solution for protanopia color deficiency, by investigating the entire spectrum on CIE L*a*b* color space. We need to find a more specific color mapping to deal with protanopia. We also want to optimize our code and develop a lightweight version that can installed on wearable devices (glasses at first) aiming to assess how comfortable and ecological it could be having an enhanced visual experience in everyday life. We also provided a new public dataset under the name of EToCVD gathering the real fixation point maps for both normal and CVD affected people involved during our experimental sessions.

Bibliography

Ala Aboudib, Vincent Gripon, and Gilles Coppin. A model of bottom-up visual attention using cortical magnification. In *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, pages 1493–1497. IEEE, 2015.

Steve Ackerman, Kathleen Strabala, Paul Menzel, Richard Frey, Chris Moeller, Liam Gumley, Bryan Baum, Crystal Schaaf, and George Riggs. Discriminating clear-sky from cloud with MODIS algorithm theoretical basis document (MOD35). *ATBD Ref. ATBD-MOD-06, version*, 4:115p, 1997. MODIS Cloud Mask Team, Cooperative Institute for Meteorological Satellite Studies, University of Wisconsin.

Turkia Al-Moustafa, Richard P Armitage, and F Mark Danson. Mapping fuel moisture content in upland vegetation using airborne hyperspectral imagery. *Remote sensing of Environment*, 127:74–83, 2012.

Mohammed Q Alkhatib, Sergio D Cabrera, and Thomas E Gill. Automated detection of dust clouds and sources in NOAA-AVHRR satellite imagery. In *Image Analysis and Interpretation (SSIAI), 2012 IEEE Southwest Symposium on*, pages 97–100. IEEE, 2012.

Zhenyu An and Zhenwei Shi. Scene learning for cloud detection on remote-sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(8):4206–4222, 2015.

Edoardo Ardizzone, Alessandro Bruno, and Giuseppe Mazzola. Visual saliency by keypoints distribution analysis. In *International Conference on Image Analysis and Processing*, pages 691–699. Springer, 2011.

Edoardo Ardizzone, Alessandro Bruno, and Giuseppe Mazzola. Saliency based image cropping. In *International Conference on Image Analysis and Processing*, pages 773–782. Springer, 2013a.

Edoardo Ardizzone, Alessandro Bruno, and Giuseppe Mazzola. Scale detection via keypoint density maps in regular or near-regular textures. *Pattern Recognition Letters*, 34(16):2071–2078, 2013b.

Edoardo Ardizzone, Alessandro Bruno, and Francesco Gugliuzza. Exploiting Visual Saliency Algorithms for Object-Based Attention: A New Color and Scale-Based Approach. In *International Conference on Image Analysis and Processing*, pages 191–201. Springer, 2017a.

Edoardo Ardizzone, Alessandro Bruno, and Francesco Gugliuzza. Eye-Tracking Through Objects (ETTO) Dataset, 2017b. URL `https://github.com/fgugliuzza/saliency/tree/master/etto`.

Caglar Aytekin, Alexandros Iosifidis, and Moncef Gabbouj. Probabilistic saliency estimation. *Pattern Recognition*, 74(Supplement C):359 – 372, 2018. ISSN 0031-3203. doi: https://doi.org/10.1016/j.patcog.2017.09.023. URL `http://www.sciencedirect.com/science/article/pii/S0031320317303734`.

Ting Bai, Deren Li, Kaimin Sun, Yepei Chen, and Wenzhuo Li. Cloud detection for high-resolution satellite imagery using machine learning and multi-feature fusion. *Remote Sensing*, 8(9):715, 2016.

Jawadul H Bappy, Sujoy Paul, and Amit K Roy-Chowdhury. Online adaptation for joint scene and object classification. In *European Conference on Computer Vision*, pages 227–243. Springer, 2016.

James C Barnes and Michael D Smallwood. Tiros-N series direct readout services users guide. 1982.

Noureddine Benabadji, Abdelatif Hassini, and Ahmed H Belbachir. Hardware and software consideration to use NOAA images. *Revue Internationale des Energies Renouvelables, CDER*, 7(01):1–11, 2004.

Nima Bigdely-Shamlo, Andrey Vankov, Rey R Ramirez, and Scott Makeig. Brain activity-based image classification from rapid serial visual presentation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 16(5):432–441, 2008.

David L Bimler, John Kirkland, and Kimberly A Jameson. Quantifying variations in personal color spaces: Are there sex differences in color vision? *Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur*, 29(2): 128–134, 2004.

Tanja Blascheck, Kuno Kurzhals, Michael Raschke, Michael Burch, Daniel Weiskopf, and Thomas Ertl. State-of-the-art of visualization for eye tracking data. In *Proceedings of EuroVis*, volume 2014, 2014.

Ali Borji and Laurent Itti. CAT2000: A Large Scale Fixation Dataset for Boosting Saliency Research. *CVPR 2015 workshop on "Future of Datasets"*, 2015. arXiv preprint arXiv:1505.03581.

Carlos Bosquez, Adrian Ramos, and Linda Noboa. System for receiving NOAA meteorological satellite images using software defined radio. In *ANDESCON, 2016 IEEE*, pages 1–4. IEEE, 2016.

Hans Brettel, Françoise Viénot, and John D Mollon. Computerized simulation of color appearance for dichromats. *JOSA A*, 14(10):2647–2655, 1997.

Bruno, A. and Gugliuzza, F. and Ardizzone, E. and Giunta, C. and Pirrone, R. Eye-Tracking of Colour Vision Deficiencies, 2018. URL `https://github.com/fgugliuzza/saliency/tree/master/etocvd`.

Manuel Bülo. A Simple GnuRadio HRPT Decoder, 2018. URL `https://tynet.eu/hrpt-decoder`.

Zoya Bylinskii, Phillip Isola, Constance Bainbridge, Antonio Torralba, and Aude Oliva. Intrinsic and Extrinsic Effects on Image Memorability. *Vision research*, 116:165–178, 2015a.

Zoya Bylinskii, Tilke Judd, Ali Borji, Laurent Itti, Frédo Durand, Aude Oliva, and Antonio Torralba. MIT saliency benchmark, 2015b.

Zoya Bylinskii, Tilke Judd, Aude Oliva, Antonio Torralba, and Frédo Durand. What do different evaluation metrics tell us about saliency models? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018.

Gustavo Camps-Valls, Jon Atli Benediktsson, Lorenzo Bruzzone, and Jocelyn Chanussot. Introduction to the issue on advances in remote sensing image processing. *IEEE Journal of Selected Topics in Signal Processing*, 5(3):365–369, 2011.

Souradeep Chakraborty and Pabitra Mitra. A dense subgraph based algorithm for compact salient image region detection. *Computer Vision and Image Understanding*, 145:1–14, 2016.

A Chandran and J Christy. A Survey of Cloud Detection Techniques For Satellite Images. *International Research Journal of Engineering and Technology*, pages 2486–2490, 2015.

Kai-Yueh Chang, Tyng-Luh Liu, Hwann-Tzong Chen, and Shang-Hong Lai. Fusing generic objectness and visual saliency for salient object detection. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 914–921. IEEE, 2011.

Li-Qun Chen, Xing Xie, Xin Fan, Wei-Ying Ma, Hong-Jiang Zhang, and He-Qin Zhou. A visual attention model for adapting images on small displays. *Multimedia Systems*, 9(4):353–364, 2003.

Liang-Chieh Chen, Yi Yang, Jiang Wang, Wei Xu, and Alan L Yuille. Attention to scale: Scale-aware semantic image segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3640–3649, 2016a.

Yung-Sheng Chen, Chao-Yan Zhou, and Long-Yun Li. Perceiving stroke information from color-blindness images. In *Systems, Man, and Cybernetics (SMC), 2016 IEEE International Conference on*, pages 000070–000073. IEEE, 2016b.

Ming-Ming Cheng, Niloy J Mitra, Xiaolei Huang, Philip HS Torr, and Shi-Min Hu. Global contrast based salient region detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(3):569–582, 2015.

M. Cornia, L. Baraldi, G. Serra, and R. Cucchiara. Predicting Human Eye Fixations via an LSTM-based Saliency Attentive Model, 2017. URL `https://github.com/marcellacornia/sam`.

M. Cornia, L. Baraldi, G. Serra, and R. Cucchiara. Predicting Human Eye Fixations via an LSTM-Based Saliency Attentive Model. *IEEE Transactions on Image Processing*, 27(10):5142–5154, Oct 2018. ISSN 1057-7149. doi: 10.1109/TIP.2018.2851672.

Susan A Cotter, David Y Lee, and Alan L French. Evaluation of a new color vision test: "color vision testing made easy". *Optometry and Vision Science: Official Publication of the American Academy of Optometry*, 76(9):631–636, 1999.

Alexandru Csete. NOAA Weather Satellite Reception with GNU Radio and USRP, 2010. URL `http://oz9aec.net/radios/gnu-radio/noaa-weather-satellite-reception-with-gnu-radio-and-usrp`.

Dartcom Systems Ltd. HRPT/AHRPT System, 2018. URL `https://www.dartcom.co.uk/products/hrpt-ahrpt-system`.

Abhishek Das, Harsh Agrawal, Larry Zitnick, Devi Parikh, and Dhruv Batra. Human Attention in Visual Question Answering: Do Humans and Deep Networks Look at the Same Regions? *Computer Vision and Image Understanding*, 163(Supplement C):90 – 100, 2017. ISSN 1077-3142. doi: https:

//doi.org/10.1016/j.cviu.2017.10.001. URL `http://www.sciencedirect.com/science/article/pii/S1077314217301649`. Language in Vision.

John D'Errico. polyfitn, 2012. URL `https://www.mathworks.com/matlabcentral/fileexchange/34765-polyfitn`.

Wenhui Diao, Xian Sun, Xinwei Zheng, Fangzheng Dou, Hongqi Wang, and Kun Fu. Efficient saliency-based object detection in remote sensing images using deep belief networks. *IEEE Geoscience and Remote Sensing Letters*, 13(2):137–141, 2016.

Chao Dong, Jinghong Liu, and Fang Xu. Ship detection in optical remote sensing images based on saliency and a rotation-invariant descriptor. *Remote Sensing*, 10(3):400, 2018.

Andrew T Duchowski. A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, & Computers*, 34(4):455–470, 2002.

Andrew T Duchowski. Eye tracking techniques. In *Eye Tracking Methodology*, pages 49–57. Springer, 2017.

K Duncan and S Sarkar. Saliency in images and video: a brief survey. *IET Computer Vision*, 6(6):514–523, 2012.

Krista A Ehinger, Barbara Hidalgo-Sotelo, Antonio Torralba, and Aude Oliva. Modelling search for people in 900 scenes: A combined source model of eye guidance. *Visual cognition*, 17(6-7):945–978, 2009.

EnChroma Inc. Enchroma, 2010. URL `http://enchroma.com`.

Stephen Engel, Xuemei Zhang, and Brian Wandell. Colour tuning in human visual cortex measured with functional magnetic resonance imaging. *Nature*, 388 (6637):68–71, 1997.

EUMETSAT. TD 18 Metop Direct Readout AHRPT Technical Description. 2013.

Dean Farnsworth. *The Farnsworth-Munsell 100-hue test: For the examination of color discrimination.* Macbeth, 1957.

John Farserotu and Ramjee Prasad. A survey of future broadband multimedia satellite systems, issues and trends. *IEEE Communications Magazine*, 38(6): 128–133, 2000.

HN Feidas, C Cartalis, and AP Cracknell. Use of Meteosat imagery to define clouds linked with floods in Greece. *International Journal of Remote Sensing*, 21(5):1047–1072, 2000.

Steve Foga, Pat L Scaramuzza, Song Guo, Zhe Zhu, Ronald D Dilley, Tim Beckmann, Gail L Schmidt, John L Dwyer, M Joseph Hughes, and Brady Laue. Cloud detection algorithm comparison and validation for operational Landsat data products. *Remote sensing of environment*, 194:379–390, 2017.

Antonino Furnari, Giovanni Maria Farinella, and Sebastiano Battiato. An experimental analysis of saliency detection with respect to three saliency levels. In *European Conference on Computer Vision*, pages 806–821. Springer, 2014.

Orazio Gambino, Ester Minafo, Roberto Pirrone, and Edoardo Ardizzone. A tunable digital Ishihara plate for pre-school aged children. In *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*, pages 5628–5631. IEEE, 2016.

Dashan Gao and Nuno Vasconcelos. Discriminant saliency for visual recognition from cluttered scenes. In *Advances in neural information processing systems*, pages 481–488, 2005.

Vito Gentile, Salvatore Sorce, Alessio Malizia, Dario Pirrello, and Antonio Gentile. Touchless Interfaces For Public Displays: Can We Deliver Interface Designers From Introducing Artificial Push Button Gestures? In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, AVI '16, pages 40–43, New York, NY, USA, 2016. ACM. ISBN 978-1-4503-4131-8. doi: 10.1145/2909132.2909282. URL `http://doi.acm.org/10.1145/2909132.2909282`.

Agostino Gibaldi, Mauricio Vanegas, Peter J. Bex, and Guido Maiello. Matlab Toolbox EyeX, 2016. URL `https://sourceforge.net/p/matlabtoolboxeyex/wiki/Home`.

Agostino Gibaldi, Mauricio Vanegas, Peter J. Bex, and Guido Maiello. Evaluation of the Tobii EyeX Eye tracking controller and Matlab toolkit for research. *Behavior Research Methods*, 49(3):923–946, Jun 2017. ISSN 1554-3528. doi: 10.3758/s13428-016-0762-9. URL `https://doi.org/10.3758/s13428-016-0762-9`.

Albano González, Juan C Pérez, Jonathan Muñoz, Zebensui Méndez, and Montserrat Armas. Watershed image segmentation and cloud classification from multispectral MSG–SEVIRI imagery. *Advances in Space Research*, 49(1):135–142, 2012.

Mark Griffiths. Turnstile design | DigitalHam, 2014. URL `http://www.digitalham.co.uk/design`.

Group for Earth Observation. APT Weather Satellite Reception - A Guide for Beginners, 2003. URL `http://www.geo-web.org.uk/apt.php`.

Jonathan Harel. A Saliency Implementation in MATLAB, 2012. URL `http://www.vision.caltech.edu/~harel/share/gbvs.php`.

Jonathan Harel, Christof Koch, Pietro Perona, et al. Graph-based visual saliency. In *NIPS*, volume 1, page 5, 2006.

Andrew K Heidinger, Amato T Evan, Michael J Foster, and Andi Walther. A naive Bayesian cloud-detection scheme derived from CALIPSO and applied within PATMOS-x. *Journal of Applied Meteorology and Climatology*, 51(6):1129–1144, 2012.

Guillaume Hervet, Katherine Guérard, Sébastien Tremblay, and Mohamed Saber Chtourou. Is banner blindness genuine? Eye tracking internet text advertising. *Applied cognitive psychology*, 25(5):708–716, 2011.

Weilong Hou, Xinbo Gao, Dacheng Tao, and Xuelong Li. Visual saliency detection using information divergence. *Pattern Recognition*, 46(10):2658–2669, 2013.

Fan Hu, Gui-Song Xia, Jingwen Hu, and Liangpei Zhang. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. *Remote Sensing*, 7(11):14680–14707, 2015.

Jia-Bin Huang, Chu-Song Chen, Tzu-Cheng Jen, and Sheng-Jyh Wang. Image recolorization for the colorblind. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, pages 1161–1164. IEEE, 2009.

Manabu Ichikawa, Kiyoshi Tanaka, Shoji Kondo, Koji Hiroshima, Kazuo Ichikawa, Shoko Tanabe, and Kiichiro Fukami. Web-page color modification for barrier-free color vision with genetic algorithm. In *Genetic and Evolutionary Computation Conference*, pages 2134–2146. Springer, 2003.

Manabu Ichikawa, Kiyoshi Tanaka, Shoji Kondo, Koji Hiroshima, Kazuo Ichikawa, Shoko Tanabe, and Kiichiro Fukami. Preliminary study on color modification for still images to realize barrier-free color vision. In *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, volume 1, pages 36–41. IEEE, 2004.

Shinobu Ishihara. *Tests for colour-blindness.* Kanehara Shuppan Company, 1960.

Laurent Itti, Christof Koch, and Ernst Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, 1998.

Jae-Yun Jeong, Hyun-Ji Kim, Young-Hyun Kim, Tae-Shick Wang, and Sung-Jea Ko. Enhanced re-coloring method with an information preserving property for color-blind person. In *Consumer Electronics (ICCE), 2012 IEEE International Conference on*, pages 600–601. IEEE, 2012.

Ming Jiang, Juan Xu, and Qi Zhao. Saliency in Crowd. In *ECCV*. IEEE, 2014.

Matthew Johnson-Roberson, Mitch Bryson, Bertrand Douillard, Oscar Pizarro, and Stefan B Williams. Discovering salient regions on 3D photo-textured maps: Crowdsourcing interaction data from multitouch smartphones and tablets. *Computer Vision and Image Understanding*, 131:28–41, 2015.

Tilke Judd, Krista Ehinger, Frédo Durand, and Antonio Torralba. Learning to predict where humans look. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 2106–2113. IEEE, 2009a.

Tilke Judd, Krista Ehinger, Frédo Durand, and Antonio Torralba. Learning to Predict Where Humans Look, 2009b. URL `http://people.csail.mit.edu/tjudd/WherePeopleLook/index.html`.

Christopher Kanan, Mathew H Tong, Lingyun Zhang, and Garrison W Cottrell. SUN: Top-down saliency using natural statistics. *Visual cognition*, 17(6-7):979–1003, 2009.

K-G Karlsson, Erik Johansson, and Abhay Devasthale. Advancing the uncertainty characterisation of cloud masking in passive satellite imagery: Probabilistic formulations for NOAA AVHRR data. *Remote Sensing of Environment*, 158:126–139, 2015.

Kyle Keen. Rtl_fm Guide: Updates for rtl_fm overhaul, 2013. URL `http://kmkeen.com/rtl-demod-guide`.

Gia Kim and Siu Man Lui. The effect of color combination on visual attention and usability of multiple line graphs. *Journal of Communications and Information Sciences*, 1:11–21, 2011.

Christof Koch and Shimon Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. In *Matters of intelligence*, pages 115–141. Springer, 1987.

Shoji Kondo. A computer simulation of anomalous color vision. *Color Vision Deficiencies*, pages 145–159, 1990.

Srinivas SS Kruthiventi, Kumar Ayush, and Radhakrishnan Venkatesh Babu. Deepfix: A fully convolutional neural network for predicting human eye fixations. *IEEE Transactions on Image Processing*, 2017.

R Lakowski. Theory and practice of colour vision testing: A review Part 1. *Occupational and Environmental Medicine*, 26(3):173–189, 1969.

M Le Goff, J-Y Tourneret, H Wendt, M Ortner, and M Spigai. Deep learning for cloud detection. In *Pattern Recognition Systems (ICPRS 2017), 8th International Conference of*, pages 1–6. IET, 2017.

Thierry Leconte. ATPDEC Home Page, 2003. URL `https://github.com/csete/aptdec`.

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521 (7553):436–444, 2015.

Guanbin Li and Yizhou Yu. Visual Saliency Detection Based on Multiscale Deep CNN Features. *IEEE Transactions on Image Processing*, 25(11):5012–5024, 2016.

Jia Li and Wen Gao. *Visual saliency computation: A machine learning perspective*, volume 8408. Springer, 2014.

Jian Li, Martin D Levine, Xiangjing An, Xin Xu, and Hangen He. Visual saliency based on scale-space analysis in the frequency domain. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(4):996–1010, 2013.

Zhenyang Li, Kirill Gavrilyuk, Efstratios Gavves, Mihir Jain, and Cees G.M. Snoek. VideoLSTM convolves, attends and flows for action recognition. *Computer Vision and Image Understanding*, 166(Supplement C):41 – 50, 2018. ISSN 1077-3142. doi: https://doi.org/10.1016/j.cviu.2017.10.011. URL `http://www.sciencedirect.com/science/article/pii/S1077314217301741`.

Maiko MI Lie, Gustavo B Borba, Hugo Vieira Neto, and Humberto R Gamba. Joint upsampling of random color distance maps for fast salient region detection. *Pattern Recognition Letters*, 2017.

Chao-Hung Lin, Bo-Yi Lin, Kuan-Yi Lee, and Yi-Chen Chen. Radiometric normalization and cloud detection of optical satellite images using invariant pixels. *ISPRS Journal of Photogrammetry and Remote Sensing*, 106:107–117, 2015.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.

David G Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

Jiebo Luo. Subject content-based intelligent cropping of digital photos. In *Multimedia and Expo, 2007 IEEE International Conference on*, pages 2218–2221. IEEE, 2007.

Richard G Lyons. *Understanding Digital Signal Processing, 3/E.* Pearson Education India, 2004.

Gustavo M Machado, Manuel M Oliveira, and Leandro AF Fernandes. A physiologically-based model for simulation of color vision deficiency. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1291–1298, 2009.

James MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297. Oakland, CA, USA, 1967.

John Magliacane. PREDICT - A Satellite Tracking/Orbital Prediction Program, 2001. URL `http://www.qsl.net/kd2bd/predict.html`.

Ali Mahdi, Mei Su, Matthew Schlesinger, and Jun Qin. A Comparison Study of Saliency Models for Fixation Prediction on Infants and Adults. *IEEE Transactions on Cognitive and Developmental Systems*, 2017.

Salman Mahmood, Muhammad Tahir Mushtaq, and Ghulam Jaffer. Cost efficient design approach for receiving the NOAA weather satellites data, 2016.

Michi Matsukura and Shaun P Vecera. Interference between object-based attention and object-based memory. *Psychonomic Bulletin & Review*, 16(3):529–536, 2009.

N. V. Kartheek Medathati, Heiko Neumann, Guillaume S. Masson, and Pierre Kornprobst. Bio-inspired computer vision: Towards a synergistic approach of artificial and biological vision. *Computer Vision and Image Understanding*, 150 (Supplement C):1 – 30, 2016. ISSN 1077-3142. doi: https://doi.org/10.1016/j.cviu.2016.04.009. URL `http://www.sciencedirect.com/science/article/pii/S1077314216300339`.

Valter Medri. Immagini captate dal Signor Ferdinando Napoleone di Palena, 2005. URL `https://digilander.libero.it/vtmedri/immagini_Napoleone.htm`.

Gary W Meyer and Donald P Greenberg. Color-defective vision and computer graphics displays. *IEEE Computer Graphics and Applications*, 8(5):28–40, 1988.

Neda Milić, Miklós Hoffmann, Tibor Tómács, Dragoljub Novaković, and Branko Milosavljević. A content-dependent naturalness-preserving daltonization method for dichromatic and anomalous trichromatic color vision deficiencies. *Journal of Imaging Science and Technology*, 59(1):10504–1, 2015.

Isaac Moradi, Huan Meng, Ralph R Ferraro, and Stephen Bilanow. Correcting geolocation errors for microwave instruments aboard NOAA satellites. *IEEE Transactions on Geoscience and Remote Sensing*, 51(6):3625–3637, 2013.

NASA/Jet Propulsion Laboratory. Space Radar Image of Lisbon, Portugal, 1999. URL `https://photojournal.jpl.nasa.gov/catalog/PIA01807`.

Jay Neitz and Maureen Neitz. The genetics of normal and defective color vision. *Vision Research*, 51(7):633–651, 2011.

Saulo A.F. Oliveira, Shara S.A. Alves, JoÃ£o P.P. Gomes, and Ajalmar R. Rocha Neto. A bi-directional evaluation-based approach for image retargeting quality assessment. *Computer Vision and Image Understanding*, 2017. ISSN 1077-3142. doi: https://doi.org/10.1016/j.cviu.2017.11.011. URL `http://www.sciencedirect.com/science/article/pii/S1077314217302035`.

Osmocom. rtl-sdr - OsmoSDR, 2012. URL `http://sdr.osmocom.org/trac/wiki/rtl-sdr`.

Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979.

Otávio AB Penatti, Keiller Nogueira, and Jefersson A dos Santos. Do deep features generalize from everyday objects to remote sensing and aerial scenes domains? In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 44–51, 2015.

Perception Data Inc. eyePilot, 2006. URL `http://www.eyepilot.com`.

Xiaoliang Qian, Junwei Han, Gong Cheng, and Lei Guo. Optimal contrast based saliency detection. *Pattern Recognition Letters*, 34(11):1270–1278, 2013.

Yahya Rahmat-Samii and Arthur C Densmore. Technology trends and challenges of antennas for satellite communication systems. *IEEE Transactions on Antennas and Propagation*, 63(4):1191–1204, 2015.

Subramanian Ramanathan, Harish Katti, Nicu Sebe, Mohan Kankanhalli, and Tat-Seng Chua. An eye fixation database for saliency detection in images. *Computer Vision–ECCV 2010*, pages 30–43, 2010.

Karl Rasche, Robert Geist, and James Westall. Detail preserving reproduction of color images for monochromats and dichromats. *IEEE Computer Graphics and Applications*, 25(3):22–30, 2005.

Ioannis Rigas, George Economou, and Spiros Fotopoulos. Efficient modeling of visual saliency based on local sparse representation and the use of hamming distance. *Computer Vision and Image Understanding*, 134:33–45, 2015.

Sourya Roy and Pabitra Mitra. Visual saliency detection: a Kalman filter based approach. *arXiv preprint arXiv:1604.04825*, 2016.

Albert Ali Salah, Ethem Alpaydin, and Lale Akarun. A selective attention-based method for visual pattern recognition with application to handwritten digit recognition and face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(3):420–425, 2002.

Alexander B Sergienko. Software-defined radio in MATLAB Simulink with RTL-SDR hardware. In *Computer Technologies in Physical and Engineering Applications (ICCTPEA), 2014 International Conference on*, pages 160–161. IEEE, 2014.

Gaurav Sharma. Color fundamentals for digital imaging. In *Digital Color Imaging Handbook*. CRC Press, 2002.

Chengyao Shen and Qi Zhao. Webpage saliency. In *European Conference on Computer Vision*, pages 33–46. Springer, 2014.

Mengyun Shi, Fengying Xie, Yue Zi, and Jihao Yin. Cloud detection of remote sensing images by deep learning. In *Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International*, pages 701–704. IEEE, 2016.

James Michael Shima. FM demodulation using a digital radio and digital signal processing. Master's thesis, University of Florida, 1995.

James J Simpson and Jason I Gobat. Improved cloud detection for daytime AVHRR scenes over land. *Remote Sensing of Environment*, 55(1):21–49, 1996.

MP Simunovic. Colour vision deficiency. *Eye*, 24(5):747, 2010.

Robert Snowden, Robert J Snowden, Peter Thompson, and Tom Troscianko. *Basic vision: an introduction to visual perception*. Oxford University Press, 2012.

McKay Moore Sohlberg and Catherine A Mateer. *Introduction to cognitive rehabilitation: Theory and practice*. Guilford Press, 1989.

MB Sruthi, M Abirami, Akhil Manikkoth, R Gandhiraj, and KP Soman. Low cost digital transceiver design for Software Defined Radio using RTL-SDR. In *Automation, Computing, Communication, Control and Compressed Sensing (iMac4s), 2013 International Multi-Conference on*, pages 852–855. IEEE, 2013.

Yusuke Sugano and Andreas Bulling. Self-calibrating head-mounted eye trackers using egocentric visual saliency. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, pages 363–372. ACM, 2015.

Lang Sun, Yan Tang, and Hong Zhang. Visual saliency detection based on multiscale and multi-channel mean. *Multimedia Tools and Applications*, 75(1):667–684, 2016.

Veronica Sundstedt, Alan Chalmers, Kirsten Cater, and Kurt Debattista. Top-Down Visual Attention for Efficient Rendering of Task Related Scenes. In *VMV*, pages 209–216, 2004.

T.F Syeda-Mahmood. Detecting Perceptually Salient Texture Regions in Images. *Computer Vision and Image Understanding*, 76(1):93 – 108, 1999. ISSN

1077-3142. doi: https://doi.org/10.1006/cviu.1999.0784. URL `http://www.sciencedirect.com/science/article/pii/S1077314299907841`.

Satohiro Tajima and Kazuteru Komine. Saliency-based color accessibility. *IEEE Transactions on Image Processing*, 24(3):1115–1126, 2015.

The GNU Radio Foundation. GNU Radio, 2001. URL `https://www.gnuradio.org`.

The GNU Radio Foundation. NOAA POES HRPT receiver, 2009. URL `https://github.com/gnuradio/gnuradio/tree/master/gr-noaa`.

The GNU Radio Foundation. Vector-Optimized Library of Kernels, 2015. URL `http://libvolk.org`.

Jan Theeuwes, Sebastiaan Mathôt, and Alan Kingstone. Object-based eye movements: The eyes prefer to stay within the same object. *Attention, Perception, & Psychophysics*, 72(3):597–601, Apr 2010.

Alexander Toet. Computational versus psychophysical bottom-up image saliency: A comparative evaluation study. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(11):2131–2146, 2011.

John K Tsotsos and Albert Rothenstein. Computational models of visual attention. *Scholarpedia*, 6(1):6201, 2011.

Zhigang Tu, Zuwei Guo, Wei Xie, Mengjia Yan, Remco C Veltkamp, Baoxin Li, and Junsong Yuan. Fusing disparate object signatures for salient object detection in video. *Pattern Recognition*, 72:285–299, 2017.

Boonyarit Uengtrakul and Dahmmaet Bunnjaweht. A cost efficient software defined radio receiver for demonstrating concepts in communication and signal processing using Python and RTL-SDR, 2014.

Françoise Viénot, Hans Brettel, and John D Mollon. Digital video colourmaps for checking the legibility of displays by dichromats. *Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain),* *Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur*, 24(4):243–252, 1999.

Eleonora Vig, Michael Dorr, and David Cox. eDN-saliency, 2014a. URL `https://github.com/coxlab/edn-cvpr2014`.

Eleonora Vig, Michael Dorr, and David Cox. Large-scale optimization of hierarchical features for saliency prediction in natural images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2798–2805, 2014b.

Fredrik Viksten, Per-Erik Forssén, Bjorn Johansson, and Anders Moe. Object Pose Estimation Database, 2009a. URL `http://www.cvl.isy.liu.se/research/objrec/posedb/`.

Fredrik Viksten, Per-Erik Forssén, Bjorn Johansson, and Anders Moe. Comparison of local image descriptors for full 6 degree-of-freedom pose estimation. In *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*, pages 2779–2786. IEEE, 2009b.

AJ Vingrys and BL Cole. Validation of the Holmes-Wright Lanterns for testing colour vision. *Ophthalmic and Physiological Optics*, 3(2):137–152, 1983.

Jeff Wallach. User's Guide for Building and Operating Environmental Satellite Receiving Stations. *National Environmental Satellite, Data, and Information Service, NOAA*, 1997.

Jan Walraven and Johan W Alferdinck. Color displays for the color blind. In *Color and Imaging Conference*, volume 1997, pages 17–22. Society for Imaging Science and Technology, 1997.

Dirk Walther and Christof Koch. Modeling attention to salient proto-objects. *Neural networks*, 19(9):1395–1407, 2006.

Hui-li Wang, Ming Zhu, Chun-bo Lin, and Dian-bing Chen. Ship detection in optical remote sensing image based on visual saliency and adaboost classifier. *Optoelectronics Letters*, 13(2):151–155, 2017.

Lei Wang, Shui-Long Dong, Hai-Sheng Li, and Xiao-Bin Zhu. A Brief Survey of Low-Level Saliency Detection. In *Information System and Artificial Intelligence (ISAI), 2016 International Conference on*, pages 590–593. IEEE, 2016.

Qi Wang, Pingkun Yan, Yuan Yuan, and Xuelong Li. Multi-spectral saliency detection. *Pattern Recognition Letters*, 34(1):34–41, 2013a.

Wei Wang, Cheng Chen, Yizhou Wang, Tingting Jiang, Fang Fang, and Yuan Yao. Simulating human saccadic scanpaths on natural images. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 441–448. IEEE, 2011.

Xin Wang, Qi Lv, Bin Wang, and Liming Zhang. Airport detection in remote sensing images: a method based on saliency map. *Cognitive neurodynamics*, 7 (2):143–154, 2013b.

Joey Wilson, Andrew Nelson, and Behrouz Farhang-Boroujeny. Parameter derivation of type-2 discrete-time phase-locked loops containing feedback delays. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 56(12):886–890, 2009.

Jimei Yang and Ming-Hsuan Yang. Top-down visual saliency via joint CRF and dictionary learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(3):576–588, 2017.

Xiwen Yao, Junwei Han, Lei Guo, Shuhui Bu, and Zhenbao Liu. A coarse-to-fine model for airport detection from remote sensing images using target-oriented visual saliency and CRF. *Neurocomputing*, 164:162–172, 2015.

Stephan R Yhann and James J Simpson. Application of neural networks to AVHRR cloud segmentation. *IEEE transactions on geoscience and remote sensing*, 33(3):590–604, 1995.

Jin-Gang Yu, Gui-Song Xia, Changxin Gao, and Ashok Samal. A computational model for object-based visual saliency: Spreading attention along Gestalt cues. *IEEE Transactions on Multimedia*, 18(2):273–286, 2016.

Yi Yuan and Xiangyun Hu. Bag-of-words and object-based classification for cloud extraction from satellite imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(8):4197–4205, 2015.

Jun Zhang and Jinglu Hu. Image segmentation based on 2D Otsu method with histogram analysis. In *Computer Science and Software Engineering, 2008 International Conference on*, volume 6, pages 105–108. IEEE, 2008.

Libao Zhang and Yingying Zhang. Airport detection and aircraft recognition based on two-layer saliency model in high spatial resolution remote-sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(4):1511–1524, 2017.

Libao Zhang, Bingchang Qiu, Xianchuan Yu, and Jindong Xu. Multi-scale hybrid saliency analysis for region of interest detection in very high resolution remote sensing images. *Image and Vision Computing*, 35:1–13, 2015a.

Libao Zhang, Aoxue Li, Zhongjun Zhang, and Kaina Yang. Global and local saliency analysis for the extraction of residential areas in high-spatial-resolution remote sensing image. *IEEE Transactions on Geoscience and Remote Sensing*, 54(7):3750–3763, 2016.

Lingyun Zhang, Matthew H Tong, Tim K Marks, Honghao Shan, and Garrison W Cottrell. SUN: A Bayesian framework for saliency using natural statistics. *Journal of vision*, 8(7):32–32, 2008.

Qiang Zhang, Yi Liu, Siyang Zhu, and Jungong Han. Salient object detection based on super-pixel clustering and unified low-rank representation. *Computer Vision and Image Understanding*, 2017.

Wenjian Zhang, Jianmin Xu, Chaohua Dong, and Jun Yang. China's current and future meteorological satellite systems. In *Earth Science Satellite Remote Sensing*, pages 392–413. Springer, 2006.

Xiujun Zhang, Chen Xu, Min Li, and Robert KF Teng. Study of visual saliency detection via nonlocal anisotropic diffusion equation. *Pattern Recognition*, 48 (4):1315–1327, 2015b.

Dan Zhu, Bin Wang, and Liming Zhang. Airport target detection in remote sensing images: A new method based on two-way saliency. *IEEE Geoscience and Remote Sensing Letters*, 12(5):1096–1100, 2015.