HOKKAIDO UNIVERSITY

Title	Absent Colors and their Application to Image Matching
Author(s)	田, 穎
Degree Grantor	北海道大学
Degree Name	博士(工学)
Dissertation Number	甲第15083号
Issue Date	2022-03-24
DOI	https://doi.org/10.14943/doctoral.k15083
Doc URL	https://hdl.handle.net/2115/85481
Type	doctoral thesis
File Information	TIAN_Ying.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

SSI-DT46195032

Doctoral Thesis

Absent Colors and their Application to Image Matching

Ying TIAN

January, 2022

Division of Systems Science and Informatics
Graduate School of Information Science and Technology
Hokkaido University

Doctoral Thesis
submitted to Graduate School of Information Science and Technology,
Hokkaido University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy.

Ying TIAN

Thesis Committee: Takayuki TANAKA Professor
Mineichi KUDO Professor
Satoshi KANAI Professor

Absent Colors and their Application to Image Matching*

Ying TIAN

Abstract

A novel approach called absent color indexing (ABC) is proposed in this thesis for robust image matching in similar objects and cluttered scenes. Image matching plays an important role in the field of computer science and technology; color features have been frequently utilized as a statistical measure of color distribution to analyze image similarity. In particular, color histograms are widely used for image matching because they have good characteristics for handling different challenges, such as rotation, deformation, scale variation, and occlusion. However, existing color histogram-based approaches focus on the main color. For similar images with few, but prominent, color features, mismatches may occur during the comparison process. Therefore, we are interested in how to make this type of color feature work.

The proposed approach provides a balanced method to focus on the contributions of absent or minor colors belonging to low-frequency or vacant bins in the histograms, which are realized by separating the source color histogram into apparent and absent color histograms. Among them, apparent colors are an essential element of conventional algorithms, serving as the main image color. Then, a threshold for this separation is obtained from the mean color histogram by considering the statistical significance of the absent colors, the frequency of which in the histogram bin is relatively low. Therefore, the inverting operation becomes particularly important. After inverting, we can effectively increase the proportion of absent colors by reversing the absent color histogram after separation. Finally, we evaluate a variety of similarity measures that can be combined with the proposed ABC, which exhibited robust image matching and distinguishability.

Algorithms based on color histogram matching exhibit better robustness, but the location information of the search target is missing in the process of color statistics. Furthermore, offset or drift often occurs in the image matching process. To overcome this problem, we propose a method that combines ABC with correlation filtering (CF) to improve matching accuracy. The CF algorithm calculates the filter by learning a template image as input and generating a two-dimensional, peak-centered, Gaussian-like model as output. Further, it relocates the matching position searched by ABC according to the generation filter to obtain the response map. The highest point in the response graph represents the best-matched position found using the combination of ABC and CF (ABC-CF).

*Doctoral Thesis, Division of Systems Science and Informatics, Graduate School of Information Science and Technology, Hokkaido University, SSI-DT46195032, January 20, 2022.

To improve the ABC approach, the offset or drift problem can be addressed by using only the color information. A multiple-layered (ML) structure is designed to add location information to the histogram-based matching process. The ML structure divides the image into three layers based on the principle that the color information of the central area remains unchanged, with the central location as the common base point. To combine ABC and ML structures (ABC-ML) in the image matching process, we perform ABC on the image of the corresponding layer. The similarities of each layer are weighted and summed to obtain the final result. These two improved algorithms improve the matching accuracy while maintaining the advantages of ABC.

This thesis is organized into the following chapters.

Chapter 1 introduces the related works in image matching and describes the importance of our research. Some of the challenges are included and discussed in image matching. Furthermore, the motivations and contributions of this study are described.

Chapter 2 presents the details of the novel concept of ABC based on the color histogram in image matching. We introduce the definitions of apparent and absent colors obtained by separating the original color histogram. Subsequently, the selection of color space is illustrated. Threshold h_T is defined to obtain apparent and absent color histograms using the mean color histogram. Four similarity measurements are discussed, which were each separately combined with ABC to evaluate the similarity of images. Therefore, many measurements can be combined with our proposed ABC methods. The margin discussion shows the distinguishability of ABC with different measurements on the Mondrian pattern. We further analyze the performance of ABC under different challenges, such as variations in illumination and scale as well as rotation, deformation, and occlusion.

Chapter 3 introduces our proposed method to utilize the proposed ABC in combination with CF. The scheme presented is effective for precise registration based on Fourier domain training and filtering, which can generate a sharp peak in the relevant output to precisely position the matching image. We train the reference as the first step. An optimal filter is obtained by calculating the minimum output sum of squared error (MOSSE) on inputs and outputs. Subsequently, the optimal filter is applied to the position selected from the search by pre-processing ABC to transform its response map into the Fourier domain. Finally, we convert the space to find the best matching position. We verify the ability of ABC-CF with adequate experiments.

Chapter 4 presents the proposed ABC-ML in detail, including the new concept of total color space and how to combine ABC and ML. An ML structure based on the isotonic principle to keep the center location is not changed. Because target images involve scale variation, rotation, and deformation, the image features at the center zone are largely retained, especially color features. Next, we aim to accurately position the target by dividing the image into multiple layers. Ultimately, each layer plays a role in restraining each other ' s positional relationship to obtain an optimized matching effect. We report the results of measurements of the robustness and efficiency of our proposed

ABC-ML on real-world images and open data.

Chapter 5 introduces the experimental evaluation. The experimental setup is described in detail. Parameter selection is discussed for comparison with fixed values. The matching performance of ABC, ABC-CF, and ABC-ML were compared experimentally, and the results are reported. We also discuss some aspects of the advantages and disadvantages of the proposed technique in this chapter. ABC has some notable benefits, including technical simplicity as well as invariance in rotation, distortion, and to some extent scaling. Furthermore, we compare ABC, ABC-CF, and ABC-ML to describe their respective performance characteristics.

The final chapter summarizes the main points of our research. Finally, we present the conclusion and suggest some possible avenues for future research.

Keywords: Color histograms, Absent color indexing (ABC), Apparent colors (AP) , Correlation filter (CF), Multiple-Layered structure (ML), Image matching

Contents

List of Figures

List of Tables

Chapter 1. Introduction

1.1 Research background

Image matching is a task of finding targets using images or some statistic or deterministic features extracted from images. The rapid development of computer vision methods has resulted in the availability of a wide variety of image information storage and extraction techniques. In recent years, various types of digital imaging equipment have been widely used with the development and popularization of image acquisition equipment, such as smartphones and UAVs in the consumer market, surveillance cameras in security, and industrial cameras in manufacturing. In most practical applications, these applications typically used together with image matching techniques. The measure of similarity or dissimilarity between images by using image matching techniques is a key task in image retrieval [1], image recognition [2], pedestrian detection [3], and defect detection [4]. Image matching [5] technology aims to identify a similar or a dissimilar structure or content from two or more images. In addition, image matching technology is applied to generate image mosaics and three-dimensional reconstruction.

Among color features, color histograms, a statistical measure of color distribution in images, have been widely used to describe color information. Such histograms are beneficial in that the color distribution in the template image is recorded without complicated learning processes. They feature strong robustness to object deformation and scale changes, and they provide effective statistics for utilizing discrete color distributions or histograms over a given color space. Thus, color histogram-based approaches [6, 7] can be effectively used to search for objects. In many applications, the color features often appear in the algorithms as a main or auxiliary feature. However, these approaches reduce performance in discriminating similar objects, because any histogram trades positional information of pixels for flexibility in matching.

With in-depth research and application of image matching, the accuracy of image matching directly affects the results of experiments and performance in practical applications. However, there are still some problems with using image matching, such as increasing amounts of image data, diverse shapes of targets, and large numbers of similar targets. Therefore, further research on methods the enable faster and more accurate target image recognition is required. Image matching methods are also widely applied in common consumer usage and practical applications, such as surveillance cameras as shown in Fig. 1.1. Using surveillance video data to identify and track specific target features is a key task in the computer vision field. To complete the task of recognition or tracking in existing video surveillance systems, it was necessary to watch the

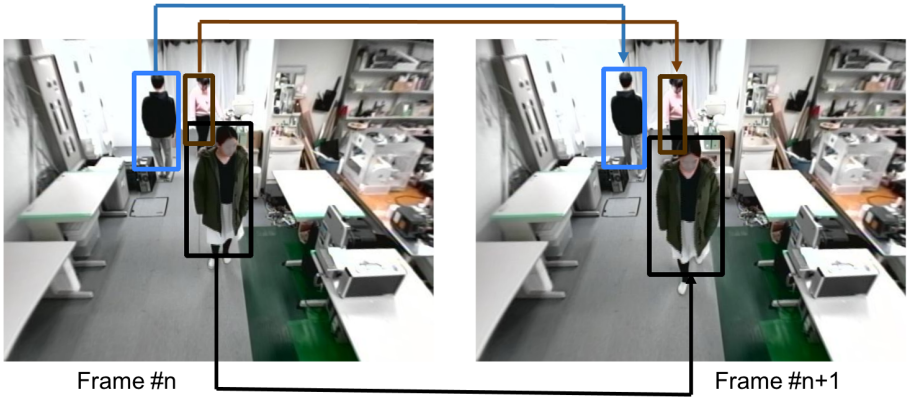


Figure 1.1: The application of image matching.

captured video content to search for targets and perform comparisons manually. This method requires considerable time to screen the video content, and the search results are dependent on the experience of human operators. Therefore, automatic extraction of target features has attracted attention as a research topic. As an essential part of this application, image matching can accurately and rapidly identify targets in a video image sequence. In the subsequent process, it can track the target by continuously matching the image and returning to its tracked position. However, the challenge is that many targets may appear with similar features. Therefore, image matching methods with discrimination abilities become significant to accurately match correct targets.

However, many image matching approaches are designed to pay attention to the prominent information of image content due to different application objectives and exhibit a lack attention directed to minor features, especially in color feature-based approaches. Therefore, methods designed to use image-matching technology to accurately judge the minor difference in images to be matched remains as an urgent problem to be solved. Therefore, while using primary features, the development of methods to provide fair treatment and analysis of minor features that play an important role would be a major breakthrough in image matching approaches based on color histograms, with applications such as of product defect detection.

In summary, image matching plays a decisive role as an important step in many applications, such as image retrieval, pedestrian detection or tracking, and defect detection technology. To better apply image matching technology in these methods, our study aims to improve accurate matching between similar images and improve the robustness of image matching in cluttered scenes.

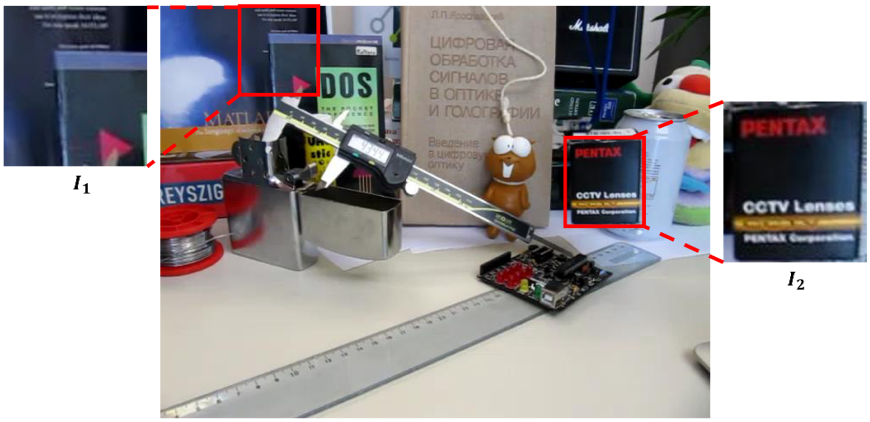


Figure 1.2: Two images I_1 and I_2 as example of same size.

1.2 Motivation of the research

By describing the background of this study, we may note that image matching faces challenges that need to be solved. An ideal evaluation method for image matching should be robust, discriminative, and stable. The ABC approach is proposed to overcome these problems. The development of the proposed approach was motivated by the aim of enhancing color-based features in cases wherein an object to be searched exhibits few, but prominent, colors and existing features. For example, eye color is a quantitatively minor or hidden color feature when identifying a person, but it can provide an essential element for identification. Especially in the cases of computational image pattern search problems, these may be somewhat hopeful to contribute separation of the targets from other candidates through enhancement of identifiers in apparent or neutral color features. In Fig. 1.2, I_1 and I_2 are examples of the same size $(100, 90)$ that exhibit remarkably similar colors; the primary color is black, which occupies a large proportion of the image, whereas yellow, red, and white are present as minor colors with relatively few pixels. The primary colors have played an essential role in existing similarity calculation methods for conventional color histogram-based matching. However, minor colors were ignored as trivial information for evaluating similar images. The proposed approach focuses on low-frequency colors in any pair of two histograms of reference and target images. Three conditional combinations concerning high and low frequencies in their respective bins are required to evaluate histogram similarity. If both two bins include high frequencies, they may have high similarities. In contrast, if only one of the two bins has a lower frequency, its contribution to the overall similarity may be much lower. The last case where they both have low frequencies has been evaluated as having only a relatively small contribution to the overall similarity. However, it is considered to be of interest in this work. This shows that the two images include the color with low frequencies represented in the bins. In our study, this case was regarded as a helpful feature in

histogram evaluation. However, we must prevent additional noise in histograms when designing algorithms because such low frequencies may be easily influenced by noise. This study aims to solve this problem by using some particular definitions of minor colors. This work provides a method of similarity evaluation from a color histogram-based approach. The separation of colors in the histogram can enable a better understanding of the composition of colors. To the best of our knowledge, the present work is the first to explore this research direction by generating minor but essential features for image matching.

1.3 Conventional image matching methods

Image matching is among the basic problems of computer vision. Such methods can search for correct images in a given scene through image information or detection and description of different features as shown in Fig. 1.3. As a result, the tasks of recognition

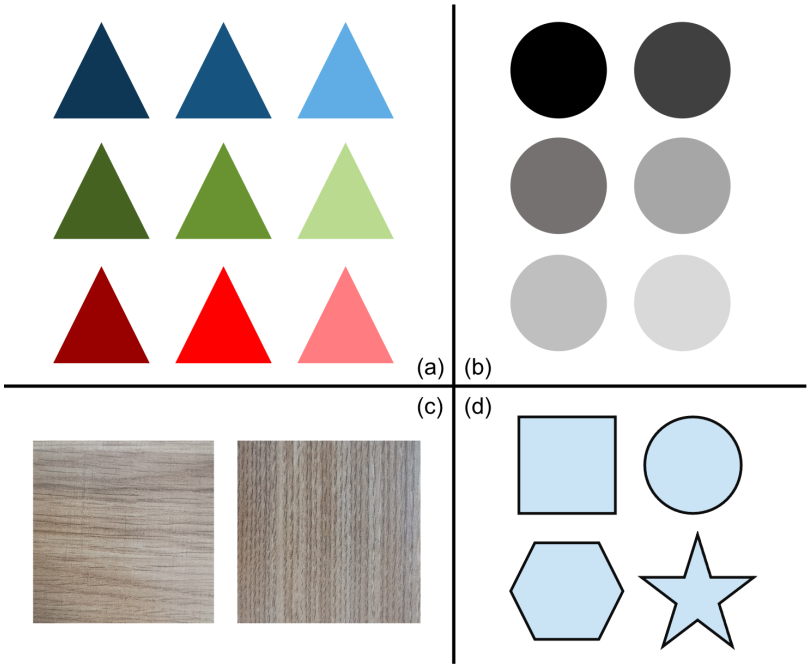


Figure 1.3: The example of different image features.

or matching are completed. Conventional image matching technology can be divided into methods based on global features and methods based on local features. Local features are features extracted from local areas of an image, including feature points, lines, edges, and areas with special attributes. The advantage of local feature matching is that the matching effect is better when the image is occluded, and the matching can be completed by the correspondence of some features. Methods based on local features also involve some shortcomings. First, the density of features extracted in image regions with less

texture is frequently sparse, making it difficult to extract local features. Second, if there are two or more repetitive features in an image, matching methods based on local features are prone to mismatches. Third, most of the feature points extracted from images with a large number of cluttered backgrounds are meaningless, leading to image matching errors. Moreover, with the increasing performance of image capture equipment, the resolution of the collected images is progressively increasing. The number of extracted features and time invested in recording them have increased greatly. As the number of feature points increases, the time required for image matching also does so. Global features are the attributes of the entire image, including color features, texture features, and shape features. Most of the global features are visual features at the pixel level of the bottom layer. Hence, they have good invariance, simple calculation, and intuitive representation. The advantage of this type of algorithm is that it can quickly extract many image features and efficiently use the context of the entire image, so it is widely used in computer vision. Image matching based on global features still has some shortcomings. For example, the matching effect is not ideal when there is occlusion or overlap in the image.

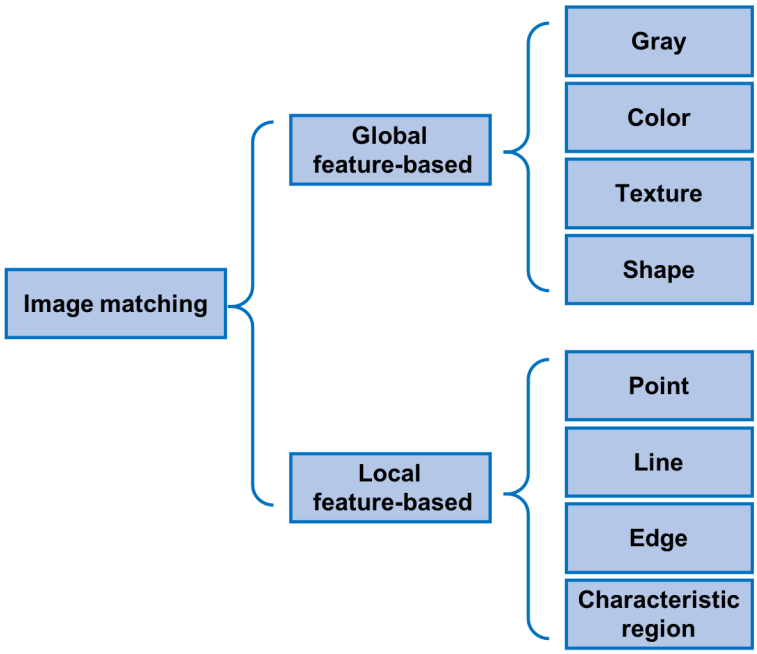


Figure 1.4: Classification of image matching methods.

The conventional image matching algorithm is widely used owing to its high operation speed and low time consumption. With the development of machine learning, such methods have been applied in the field of computer vision. An increasing number of researchers have adopted machine learning methods to improve algorithms in each stage of image matching. Image feature extraction and feature description have a stronger

ability to extract features than conventional methods. However, the methods based on machine learning require a large amount of data to perform training. In some application contexts, such as defect detection, it is typically difficult to obtain sufficient training data, such as defect detection. Hence, the development of a robust and time-saving algorithm is essential.

In the previous discussion, we introduced image matching methods based on the use of different features in the image to classify and the advantages and disadvantages of different classification methods. Fig. 1.4 shows a classification summary of image matching methods. In the following section, we introduce in detail some of the image matching methods related to this research, including color- and histogram-based matching methods.

1.3.1 Color feature-based methods

Color features are basic and direct visual features that describe the content of an image and are widely used in image matching. In addition, color features are generally stable and are not sensitive to noise, size, resolution, and direction changes. However, due to the different goals of various image matching tasks, matching a single color feature adequately may be difficult. Thus, the method of combining color features with other image features has also been widely investigated in image matching. In this section, the color features are mainly introduced, and then some other features are described that combine with color features, such as grayscale, texture, and shape features. Table 1.1 shows the commonly used feature-based matching methods.

In the image matching process, different methods should be used to extract color features after color space transformation. The extraction and description of color features is the most important part of the image matching method based on color features. The existing color features mainly include color histograms [8], cumulative color histograms [9], color co-occurrence histograms [10], fuzzy color histograms [11], dominant color descriptors [12, 13], color sets [14], color moments [9], color correlograms [15], color coherence vectors [16], and color co-occurrence matrices [17].

Color space

Before color feature extraction, it is frequently necessary to select a suitable color space. Essentially, the color space is a three-dimensional coordinate system, and a three-dimensional coordinate point can represent each color in the system. At present, the commonly used color spaces [18] mainly include RGB, HSV, and CIE L*a*b* space. The RGB color space is based on the three basic colors of R (red), G (green), and B (blue), which are superimposed in different degrees to produce a rich and wide range of colors, so it is commonly known as the three-primary color mode. The RGB mode can represent more than 16 million different colors. It is quite close to the natural colors

Table 1.1: Summarization of the feature-based matching methods.

Feature	Method
Color	Color histogram
	Cumulative color histogram
	Color co-occurrence histogram
	Fuzzy color histogram
	Dominant color descriptor
	Color sets
	Color moments
	Color coherence vector
	Color correlogram
	Color co-occurrence matrix
Grayscale	Sum of squared differences
	Normalized cross correlation
	Mutual information
Texture	Gray-level co-occurrence matrix
	Tamura
	Wavelet transform
	Gabor filter
	Curvelet transform
	Local binary pattern
	Local directional pattern
Shape	Histogram of oriented gradients
	Edge histogram descriptor
	Geometric moment invariant
	Chain code histogram

normally perceived by the human eye, so it is also called the natural color mode. The greatest advantage of the RGB color space is that it is intuitive and easy to understand. However, the disadvantage is that the three components of R, G, and B are highly correlated; if a certain component of color changes to a certain extent, then the color is likely to change. Another disadvantage of the RGB color space is the quite poor uniformity.

The HSV color model defines colors based on humans' intuitive perception of color,

based on light, and shade, and hue, where H (Hue), S (Saturation), and V (Value) represent chroma, color saturation, and lightness, respectively. This color system is considered better than the RGB system in that it is closer to human experience and perception, and thus it is widely used in computer vision.

CIE L*a*b* is composed of one luminance channel and two color channels. Each color is represented in the CIE L*a*b* color space by three numbers L*, a*, and b*, which represent brightness, the component from green to red, and the component from blue to yellow, respectively. CIE L*a*b* is designed based on human perception of color; more specifically, it is intended to be perceptually uniform. The CIE L*a*b* space involves many notable advantages. First, lightness and color are separated, the L* channel has no color information, and the a and b channels only have color. Second, the color range is wide, which includes all the color ranges of RGB and expresses colors that it does not include. It also compensates for the uneven distribution of the RGB color model.

Color histogram

Color histograms are the earliest proposed method of color features and remain widely used. Extensions such as cumulative color, color co-occurrence, and fuzzy color histograms, as well as dominant color descriptor are all considered color histograms. The study mainly adopts the method based on color histogram, so we introduce this part in detail in Sec.1.3.2.

Color moments

Color moments are another simple and effective color moment proposed by Stricker and Orengo [9]. The mathematical basis of this method is that any color distribution in the image can be represented by using the moment. In addition, because the color distribution information is mainly concentrated in the low-order moments, only the first-order moment (mean), the second-order moment (standard deviation), and the third-order moment (skewness) of the color are sufficient to express the color distribution of the image. Compared with the color histogram, another advantage of this method is that vectorizing the features is not necessary. Therefore, the color moment of the image only requires nine components (3 color components, 3 low-order moments on each element), which is more concise that those of other color features. In practical applications, to avoid the weak resolving power of low-order moments, color moments are often used in combination with different parts and generally play a role in narrowing down the color information before the use of other features. Kadir et al. [19] used color moments as color features, combined with polar Fourier transform and inter-vein pattern features to perform plant foliage image matching. The performance of this method was better than that of the Fourier transform. In a matching experiment on 50 species of plants, the accuracy rate reached 90.80%. Saikrishna [20] and others proposed a method

that divided the image into multiple parts, extracted the color moment for each region, performed clustering to calculate the average color moment, and finally used the sum of absolute differences method to determine the image matching similarity, compared with other content-based image retrieval methods available at the time, it had better performance. In a content-based image matching system based on color moment features, Mosbah et al. [21] used different color models to examine images of matching results and found that the color moment method works better in HSV color space than RGB space.

Color sets

The color set is an approximation of the color histogram. Smith and Chang proposed the color set feature to quickly find an image that matches the content of a target image in a large-scale image library [14]. First, the color image is converted into a visually balanced color space, such as an HSV space, and the color space is quantized into several bins. Then, the space is divided into several regions using the intuitive color segmentation technology. Each part is indexed by a specific color component of the quantized color space so that the image is expressed as a binary color index set. In image matching, the distance between different image color sets and the spatial relationship of color regions are compared, including the separation, inclusion, and intersection of parts, each corresponding to a different score. Because the color set is expressed as a binary feature vector, a binary search tree can be constructed to speed up the retrieval speed, which is beneficial for large-scale image collections. In the process of using color sets to represent images, relevant color combinations should be determined. Thus, Mlsna [22] and others provided improved methods of calculating related color combinations, which were able to evaluate corresponding color combinations more effectively and accelerate the color set extraction and description process.

Color-coherence vector

To address the lack of color histograms and color moments that cannot express the spatial position of the image color, Pass et al. [16] proposed the color-coherence vector of images. The color aggregation vector is also an evolution of the color histogram. The core concept is to divide the pixels belonging to each bin of the histogram into two parts. If the area of the contiguous space occupied by some pixels in the bin is more significant than a given threshold, the pixels in the area are regarded as aggregated pixels; otherwise, they are regarded as non-aggregated pixels. Because it contains the spatial information of the color distribution, the color aggregation vector can achieve better image characteristics than the color histogram. Ravani et al. [23] analyzed the performance of different stages of the color aggregation vector, optimized each step, improved the concurrent implementation of the algorithm, and used the parallel calculation method to speed up the calculation of the color aggregation vector method. Furthermore, Hamami

et al. [24] optimized the number, distance, and angle of color coherence regions. They used histogram technology to represent the information of each color coherence region and combined the number and location information of coherence regions with the color aggregation vector to improve the matching effect. Recently, Singh et al. [25] used color features based on a color aggregation vector combined with texture features based on the Gabor filter to perform content-based image retrieval to obtain the best-matching image. The accuracy of their proposed approach improved on that of popular image retrieval methods.

Color correlogram

Color correlation diagrams are an alternative expression of the color distribution of images [15]. Zhao [26] and others used different correlation maps as image features for matching, retaining spatial information, and detecting the rotation change of the target during matching. Moreover, using the gradient descent method average moving algorithm to calculate the best position and direction can accurately track rotating targets. In a target experiment, this method demonstrated the ability to accurately complete tracking tasks involving movement, size changes, or occlusions. Rasheed et al. [27] proposed an image separation technique that first decomposes the image into four bins, displays each bin, and extracts the maximum regeneration to generate a color correlation map. Their experimental results demonstrated the validity of their proposed method. Vinayak [28] and others combined a color autocorrelation map with colors when extracting color features and added Gabor filter texture features used as image features in retrieval tasks. The method of retrieving matching results is more reasonable.

Color co-occurrence matrix

The color co-occurrence matrix is a special type of color feature, which contains only color but also texture information. It uses the co-occurrence matrix method to integrate the color and texture information in the image [17]. The color co-occurrence matrix can be divided into single-channel and multi-channel color co-occurrence matrices according to the combination of color channels. Among them, the single-channel color co-occurrence matrix is only a descriptor similar to the gray-scale co-occurrence matrix extracted from a single color channel image and does not contain the spatial correlation of different channel colors. The multi-channel color co-occurrence matrix is spatially related to the colors of other channels to achieve quantification summarizes the color and texture attributes of the image [29]. Vadivel et al. [30] proposed an improved method based on the color co-occurrence matrix called an integrated color and intensity co-occurrence matrix and used this feature to perform content-based image retrieval. Compared with other color and texture-based co-occurrence matrix features, it showed a higher matching accuracy. Lin et al. [31] studied the integration of three image features

for image retirement: color or co-occurrence, the difference between pixels of scan pattern, and color histogram for K-mean. They used three features: color features, texture features, and space features. When matching, different features were selected through sequential forward selection, and the optimal feature was used for matching search. Losson et al. [32] proposed a new color feature, chromatic co-occurrence matrices in color filter array images based on color co-occurrence matrices. They compared colors and textures through the similarity of color filter array and color co-occurrence matrices. Thangarasu [33] and others combined the bit pattern feature and color co-occurrence feature with image matching based on the image ' s color, texture, and edge features, thereby enhancing the retrieval accuracy.

Others

The use of a single color feature for image matching will typically lose the information of other components in the image, which may affect the accuracy of image matching in some cases. Thus, some methods have been developed that combine multiple elements investigated in previous research using different color features for image matching. For example, the color co-occurrence matrix includes combined color features and gray-scale texture features to develop a new feature description method. Most of the multi-feature fusion matching methods can compensate for the shortcomings of different parts and improve the matching performance. Because grayscale, texture, and shape features are the same as color features, they are all widely used low-level visual features. The calculation is simple and efficient; frequently, these features are considered with color features to apply to image matching. The feature combination method can directly use the existing mature grayscale texture, shape features, and matching methods. Therefore, we introduce existing image matching methods and, in the following, describe research status of other features.

The grayscale image matching method is also called the template matching method [34]. Template matching uses a given image as a template and searches for the best-matched target image among other images. In the process of image matching, there is no need to only detect the feature information in the target, and the correlation estimation is directly performed by using the region of interest of the predefined size or even the entire image size window. In spatial domain matching, similarity measures include the sum of squared differences (SSD) [35], normalized cross-correlation (NCC) [36], and mutual information (MI) [37].

Image texture is also an important image feature that describes the inherent surface characteristics of a specific object and its relationship with the surrounding area. It reflects the spatial distribution law of the neighborhood grayscale of pixels. The most commonly used texture features in image matching mainly include gray-level co-occurrence matrices [38], Tamura textures [39], wavelet transforms [40], Gabor filters [41], curvelet transforms [42], local binary patterns [43], and local directional pattern [44]. Along

with many texture-based methods, they have been proposed and applied to image-based matched image retrieval aspects [45].

Shape features [46] are another basic visual feature that describes image content. Commonly used shape features include histogram of oriented gradients (HOG) [47], edge histogram descriptor [48], geometry geometric moment invariant (Hu moment) [49], and chain code histogram [50]. In a sense, the shape feature contains certain semantic information. When the texture and color information is insufficiently rich, the accuracy can be improved from the perspective of shape features. Shape features are robust to image displacement and scale transformation and will not be affected by different color information, but the shape expression and description are complicated. Because the three-dimensional scene is projected into a two-dimensional image and one-dimensional information is lost, the image shape feature can only partially express the scene. Therefore, the shape feature is often damaged by noise, defects, accidental deformation, and occlusion.

1.3.2 Histogram-based methods

Among image matching methods based on histograms, color histograms are the most commonly used method. The color histogram records each color that appears in an image and the ratio of each color in the image. Frequently, a rectangular coordinate system can be used to represent the results of the color histogram intuitively. The abscissa represents the color value of various colors appearing in the image, and the ordinate represents the proportion of each color in the image. As the most commonly used feature-matching method in image matching, color histograms have the following advantages. (1) They allow intuitive expression and convenient calculation. They describes the ratio of each color in an image, which has significant advantages for image matching where the color space distribution can be ignored. (2) They are not sensitive to image rotation, translation, size change, and other types of processing. The histogram calculated by this method is accurate, and the method does not yield different calculation results due to such changes in the image. (3) They are superimposable. Suppose an image is divided into several blocks. In this case, the color histogram of each block is calculated separately, and the blocks are added together to obtain the color histogram of the entire image. Although the color histogram is widely used given the above advantages, it also has many shortcomings. The greatest weakness is that this method only describes the overall ratio of various colors in the image and does not consider the spatial distribution of multiple colors. Therefore, two images with the same color histogram may have different color spatial distributions, and the resulting image contents may differ significantly. Swain et al. [51] proposed the color index method (CI), which uses color histograms for image matching. This method uses the histogram intersection to calculate the similarity between the images, thereby obtains the degree of matching images. The advantages of the CI method include its simple calculation, fast speed,

and successful matching even when the target is deformed or scaled. However, it is more sensitive to noise and brightness changes. Subsequently, a matching method was proposed based on the cumulative color histogram (CCH) [9]. The bin value of each color in CCH is the sum of all probabilities less than this color value. During image matching, the color of the image sometimes cannot be taken over all possible colors, resulting in zero values in the color histogram. A zero value in the histogram affects the intersection operation of the histogram so that the matching value cannot correctly reflect the color difference between the two images. This problem is solved based on the CCH matching method. The L1 and L2 distance between the two cumulative histograms are used as the similarity measure to perform matching. Compared with the CI method, this method has similar matching accuracy but is more robust. Chang [10] et al. used the color co-occurrence histogram (CH) as a color feature for image matching for object recognition. The color co-occurrence histogram adds geometric information to the normal color histogram so that the new feature contains both color information and geometric information. Moreover, by setting different co-occurrence distances, the algorithm can adapt to the geometric changes of the target. This method can accurately identify objects when the target is deformed and partially occluded. However, it requires high calculation time. Han et al. [11] proposed a color histogram feature called a fuzzy color histogram. It used the fuzzy set membership function to associate the color of each pixel with each bin in the color histogram through similarity to obtain the fuzzy color histogram. Using this feature for image matching can overcome small illumination changes and noise interference. After the most representative color histogram-based methods above were proposed, more color histogram matching methods were applied to different fields. Subsequently, a series of fuzzy color histogram-based methods have been proposed to overcome the problems of noise interference and illumination. Verma et al. [52] introduced triangular membership functions to improve fuzzy color histograms for template matching (TFCM).

Singh et al. [53] combined color histogram features and statistical texture features based on wavelet theory, used Euclidean distance to measure similarity, and used it in X-ray image retrieval to improve the efficiency of image analysis. Jasmine et al. [54] combined color histograms with maximum local edge binary pattern joint histograms for content-based image retrieval based on the color histogram and significantly improved the retrieval effect through the fusion of multiple features. Zeng et al. [55] used a Gaussian mixture model to quantify the color space and generate a color histogram as an image feature. The Shannon divergence was used to measure the image similarity to achieve the goal of image retrieval. The experimental results showed that their proposed approach showed the advantage of strong practicability. Gupta et al. [56] used color correlation maps to supplement the missing spatial information of color histograms for image retrieval. This method had the advantages of high retrieval efficiency and insensitivity to the location of the scene. Sun et al. [57] used statistical principles, improved the color histogram, combined it with the features of the gray-level co-occurrence matrix,

normalized them, and then merged them. This method effectively improved the original weakness of insufficient expression of color feature information. Shen et al. [58] combined the color histogram and the edge direction histogram to calculate the similarity with the improved Canberra distance, which improved the final retrieval accuracy and robustness. Soni et al. [59] added color-related features using color histograms and used Euclidean distance to calculate the similarity of matching to perform efficient image retrieval. Lacheheb et al. [60] proposed a hybrid clustering method for image retrieval. This method first extracts the mean SIFT feature of the image as a local feature while also extracting the HSV color histogram feature as a global feature. Then, local and global features are combined to maximize the segmentation of the target image and the background image, thereby ensuring the performance effect of each feature descriptor in the image description. Mohammed et al. [61] used color histograms as color features and discrete cosine transform to extract image texture features, combining the two features to perform image matching. Using Manhattan distance, Euclidean distance, and mean square error as similarity measures for analysis, the results indicate that Manhattan distance performed better for retrieval accuracy when used as similarity measures. Bhunia et al. [62] improved the color histogram, extracted the relationship between the H channel and the S channel in the HSV color space as the color feature, and proposed a diagonally symmetric local binary co-occurrence pattern method to record the texture feature description Symbiosis between adjacent textures. Finally, the retrieval result was satisfactory when combining the color and texture features with the best strategy.

Comaniciu et al. [63] proposed calculating the Bhattacharyya distance through the color histogram of the target image and the reference image and iteratively finding the minimum distance to obtain the target position and complete the target tracking. Bertinetto et al. [64] combined the correlation filter with the color histogram. They used the complementary advantages of the color histogram and the correlation filter method to address the poor tracking effect of the correlation filter method when the shape of the target changes. While improving the tracking performance, their proposed method also exhibited good real-time performance. Zhang et al. [65] proposed a multi-feature fusion tracker, combining local and global tracking methods. At a local level, a tracking method for the structural local color histogram feature block is proposed. At a global level, a correlation filter tracker based on HOG features is used. The two trackers are fused when calculating the response to achieve good tracking performance. Abdelali et al. [66] proposed a new accept–reject color histogram-based matching method, in which the similarity measure uses Bhattacharyya kernel. The target area was determined by the similarity with the target image to track the target. This method can track the target in real-time and is not sensitive to the target's rotation scale change. Possegger et al. [67] researched a target tracking method based on color representation to address tracking errors due to areas similar to the target appearing in the tracking process. An improved color histogram matching method was proposed to identify potential errors in the tracking process and improve tracking accuracy. Lukežic et al. [68] introduced

channel and spatial reliability into discriminative correlation filter tracker by using color histograms to increase the search area of target tracking and improve the tracking effect of non-rectangular targets. Fan et al. [69] used a color clustering-based histogram model to replace the color histogram model to enhance the ability to distinguish between the target and background [64] and added the structure information of the target. Then, their proposed method learn spatiotemporal regularized CFs, which improved the tracking effect of this method when the target had a large deformation. Hao et al. [70] offer a complementary tracker based on structural patch response fusion. Lightning and shape changes can be automatically processed during matching through the fusion of correlation filter and color histogram models. They carried out tracking experiments on multiple datasets, and their proposed approach has shown excellent performance.

In these tasks, the matching method based on the color histogram has gradually developed from a single feature matching to a multi-feature fusion matching method after years of research and development. The matching method of multi-feature fusion can effectively improve matching accuracy by including more image information. However, there are also some problems. First, excessive amounts of features may cause information redundancy, which increases the burden of the algorithm. Second, the method of multi-feature fusion frequently drastically increases the amount of calculation for image matching. Finally, in the matching of multiple feature fusions, the increase of information also increases the matching constraints, which may lead to a decrease in matching flexibility. In summary, more research is required on the multi-feature fusion image matching method, and an excellent single-feature image matching method is still a problem worthy of research.

1.4 Challenges and contributions in image matching

Challenges

At present, there are many highly effective image matching methods, which are used in different image processing tasks. However, many challenges remain to be solved in the image matching process. Because the varying purpose, conditions, and environment, the challenges encountered are also diverse. However, image matching involves some common challenges, as shown in Fig. 1.5, including illumination variation [71], rotation [72], deformation [73], scale variations [74], and partial occlusion [75]. Existing research also aimed to solve these challenges in the matching process effectively. The following summarizes the challenges in image matching.

- **Illumination variation**

  The problem of illumination variations is an old problem in computer vision, but a critical factor that affects the accuracy of image matching. There are many reasons for the changes in the illumination of the image. Including environmental changes, collection equipment changes, and target changes. Environmental changes, such

(a) Reference (b) Illumination variation (c) Rotation

(d) Deformation (e) Scale variation (f) Occlusion

Figure 1.5: The case of challenges in image matching.

as outdoor images, are primarily due to changes in illumination caused by natural light over time and changes in lighting due to changes in weather and environment. Environmental changes are divided into indoor and outdoor situations. The illumination variation of outdoor images is mainly caused by natural light over time and changes in the weather environment. Indoor lighting changes especially come from the opening and closing of the light source, adjustment of the brightness, and change of the position of the light source. Different types of collection equipment or changes in the location and parameters of the collection equipment also cause image illumination variation. Moreover, the movement of the target causes illumination variation. Changes in illumination lead to changes in pixel values in images and produce images of color characteristics for color images. This has a significant influence on the method of matching based on gray and color features. In addition, severe lighting changes affect the texture and contour features of the image. For example, overexposure or insufficient image brightness causes the texture features of the image to disappear or change the shape of the target. This also affects the accuracy of image matching based on texture features and shape features.

- **Deformation**

  Deformation in images is a considerable challenge in image matching methods, such as template matching, face recognition, and target recognition and tracking. The deformation in the image is mainly caused by the movement of the target or the position and angle changes when the image is collected. Many methods in template matching are matched based on the one-to-one correspondence between

pixels. This method is quite sensitive to the deformation in the image, which can easily lead to the failure of the matching. In facial recognition, the matching is mainly based on the geometric features of the human face. However, due to the change of human expression, the face will be deformed, which will cause the face recognition accuracy to decrease. Methods to recognize facial deformation and perform accurate matching have become noted difficulties of this technology. In addition, deformation is a common interference problem in target tracking. When the posture of the moving target changes, its characteristics and appearance model change, which drastically reduces the matching success rate and easily leads to tracking failure.

- **Rotation**

  There are two cases of rotation in two-dimensional image matching: the rotation of the image and the rotation of the target in the three-dimensional space. In some template matching methods, image rotation is a considerable challenge in the matching process. Frequently, more matching times are required to complete the matching, which significantly increases the amount of calculation. Moreover, achieving the matching of images not based on rotation-invariant features is typically difficult. The rotation of the target in the three-dimensional space is a significant challenge in image matching methods. This type of rotation causes the disappearance of some original features and the generation of new features, which significantly impacts the matching of grayscale, texture, contour, and color features. Therefore, the development of method to perform accurate image matching in the presence of rotation remains as a notable research problem.

- **Scale variation**

  Scale changes are frequently caused by changes in the distance between the collection device and the target. When the image size does not change, the scale change of the target to be matched will cause two situations to occur. One case is that the smaller the target leads to a substantial increase in the proportion of background in the image to be matched. Another situation is that the larger the target causes a part of the target to exceed the image range, reducing image characteristics. Both of these situations will reduce the success rate of matching. Therefore, realizing scale adaptive matching is a considerable challenge in image matching.

- **Occlusion**

  In image matching, occlusion is a severe challenge for almost all matching methods based on global features. Moreover, although the matching method based on local features can tolerate the existence of occlusion to a certain extent, it leads to a decrease in matching accuracy. However, occlusion conditions are widespread in various tasks. Especially in pedestrian tracking in a surveillance environment, an object to be matched may be blocked by hats, backpacks, and other accessories or

blocked by other pedestrians or buildings. The existence of occlusion affects the feature extraction and matching, leading to tracking failure. For image matching with occlusion, the problem is difficult to solve it in a targeted manner. The best approach is to perform a more robust feature extraction.

Contributions

In this paper, we propose a new method for image matching based on ABC. Our proposed approach performs robust matching under illumination, rotation, deformation, occlusion, and scale variation. The contribution of this thesis are summarized as follows:

- We design an approach called ABC by enhancing the performance of minor or non-existing colors. ABC uses a decomposition method to separate color histograms into apparent and absent color histograms to perform image matching. The proposed method is focused on both the apparent and absent colors.
- The ABC approach can be robust under illumination variation via the use of the CIE L*a*b* color space to generate a two-dimensional color histogram. Among them, L-channel as illumination information is ignored to realize the property of robustness.
- A stable threshold definition is achieved. To distinguish between the use of a fixed threshold, the ABC approach obtains a mean color histogram from two compared images, where the mean color histogram represents the trend of colors between the compared images. It can effectively define the absent colors and find a suitable threshold to decompose the histogram, which is of great significance to this study.
- Sensitivity to minor or non-existing colors in the target image can be realized by inverting absent color histograms. After separating the colors, the absent colors in the color histogram have a rather lower frequency in the absent color histogram. Inverting the information is a vital operation for emphasizing the importance of absent colors. Meanwhile, apparent colors as a part of the ABC approach also conduct similarity evaluation jointly with absent colors by similarity measures.
- ABC-CF is introduced to improve the demerits of color histogram-based methods. We select the CF method for the relocated matching position after ABC searching. ABC can maintain the performance of margin, and then CF can solve the offset or drift problem in the process of matching. Compared with some machine learning approaches, the proposed ABC-CF is simpler and involves a lower training cost.
- An image matching method based on a ML structure has been also proposed, which only uses color information during the process. We train the total color space for a certain range of colors in CIELAB color space. Then, total color space is utilized to generate the color histogram. Subsequently, based on the principle that the

central feature remains unchanged, the image is structured and layered. ABC-ML used to perform more accurate matching to improve the problem of offset or drift.

1.5 Overview

The remainder of the paper is organized as follows.

In Chapter 2, ABC is introduced based on the color histogram in image matching. The definitions of apparent and absent colors are obtained by separating the initial color histogram. The selection of color space is illustrated. Subsequently, we define threshold h_T to obtain apparent and absent color histograms by using the mean color histogram. Four similarity measurements are discussed to be combined with ABC to evaluate the similarity of images. We verify that ABC can combine with many similarity measurements as a feature. The margin discussion shows the distinguishability of ABC with different measurements on the Mondrian pattern. Various challenges, such as illumination variation, rotation, deformation, scale variation, and occlusion, are designed to analyze the performance of ABC.

In Chapter 3, we introduce the concept of ABC combined with CF. This is an effective scheme for precise matching based on training and filtering in the Fourier domain, which can produce sharp peaks in the output to achieve the accurate localization of matched images. The reference image is trained by the affine transformation to obtain input samples. The minimum output sum of squared error (MOSSE) on inputs and outputs is utilized, and then an optimal filter is generated. Thereafter, the filter is applied to the searched position by pre-processing ABC to obtain its response map in the Fourier domain. Finally, we convert the space from the Fourier domain to the image domain to find the best-matching position. In the experiments, we verify the ability of ABC-CF with adequate experiments.

In Chapter 4, the design of ABC-ML is provided in detail, including the new concept of total color space and how to combine ABC and ML. The ML structure based on the isotonic principle retains the center location immutably. Because target images include scale variation, rotation, and deformation, the image features at the center zone are essentially retained, especially color features. Then, we aim to accurately position the target by dividing the image into multiple layers. Finally, each layer plays a role in restraining one another ' s positional relationship to obtain the optimized matching effect. We measure the robustness and efficiency of our proposed ABC-ML in real-world images and open data.

In Chapter 5, we introduce the experimental evaluation on image matching tasks. We describe the experimental setup in detail. Parameter selection is discussed for comparison with the fixed value. The performance of ABC, ABC-CF, and ABC-ML is compared in the matching experiment. We also discuss some aspects of the advantages and drawbacks of the proposed technique in this chapter. The merits of ABC include its technical simplicity and invariance to rotation, distortion, and to some extent scaling. Finally, we

compare ABC, ABC-CF, and ABC-ML to describe the merits and demerits.

In Chapter 6, we conclude the whole dissertation and present some possible avenues for future research.

Chapter 2. Absent color indexing (ABC)

2.1 Definition of apparent and absent colors

By introducing absent colors to realize this idea, we propose a novel approach to utilizing color histograms for robust pattern identification. This approach focuses on low-frequency colors in any two histograms of the target image. When evaluating histogram similarity, there must be four conditional combinations with respect to high and low frequencies in their bins. If both bins include high frequencies, they have high similarity, and if one is low and the other is high, it provides a low contribution to the total similarity. The case where both have low frequencies is conventionally evaluated as having a low contribution to similarity, but recognizing this as a common characteristic in our trials formalizes their treatment in similarity evaluations as an effective new approach. However, we must prevent contamination by additional noise in histograms when designing algorithms, because noise may easily influence such low frequencies. Table 2.1 defines histograms H and G in a same color space or specification, which

Table 2.1: Combinations of apparent (AP) and absent (AB) colors.

Histograms		H		
		h^{AP}	h^{AB}	0
G	g^{AP}	$(h^{\mathrm{AP}}, g^{\mathrm{AP}})$	$(h^{\mathrm{AB}}, g^{\mathrm{AP}})$	$(0, g^{\mathrm{AP}})$
	g^{AB}	$(h^{\mathrm{AP}}, g^{\mathrm{AB}})$	$(h^{\mathrm{AB}}, g^{\mathrm{AB}})$	$(0, g^{\mathrm{AB}})$
	0	$(h^{\mathrm{AP}}, 0)$	$(h^{\mathrm{AB}}, 0)$	

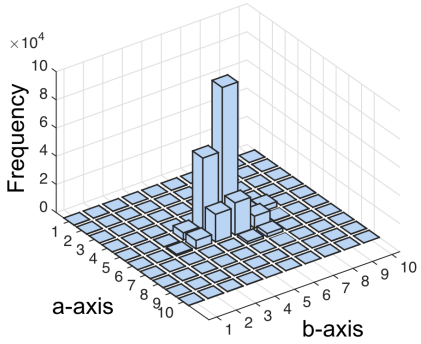
consist of high-frequency bins h^{AP} and g^{AP}, low-frequency bins h^{AB} and g^{AB}, and bins with no entities 0 and 0. Below, we will propose a detailed scheme for defining these by use of a reasonable threshold value. As Table 2.1 shows, we define *apparent colors* (AP) as those in high-frequency bins, while colors in low- and null-frequency bins provide candidate *absent colors* (AB). There are nine possible arrangements of bins having the same colors, as shown in Table 2.1. In the top row and the leftmost column, every pair includes h^{AP} or g^{AP}, which are evaluated for similarity in almost all methods based on color histograms as features; if both bins are AP their contribution to similarity may be larger, while those including an AB element make a lower contribution. As mentioned in

the previous section, our motivation is to focus on other items in the table, namely those such as $(h^{\mathrm{AB}}, g^{\mathrm{AB}})$ where both elements are AB. Such pairs are conventionally evaluated as minor elements in similarity calculations. Note that the last possible combination, 0 and 0, is never evaluated in any similarity evaluation scheme, because they represent colors not present in target images and thus are beyond the scope of consideration.

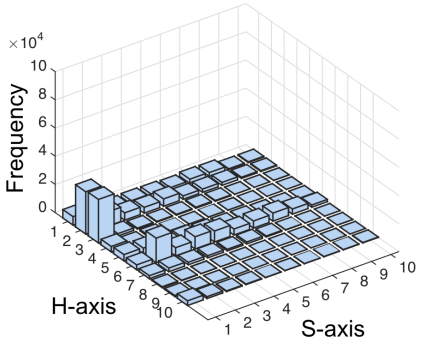
2.2 Color space selection

Many color spaces have been proposed in color management and image processing, e.g., HSV, YUV, and CIE L*a*b*. When establishing many color features in color statistics, the CIE L*a*b* color space has a broader color gamut that is closer to human vision, allowing separation brightness as an independent coordinate (the L* channel). In this study, the CIE L*a*b* color space [76] was used for ABC because it is a perceptually uniform space, where the color distribution shows a concentrated distribution trend, and the L* channel expresses the lightness. As shown in Fig. 2.1, RGB image as reference is used to generate color histogram in different color spaces. In CIE L*a*b* color space, the concentrated distribution is verified which the bins are in the center of the histogram. On the contrary, there is a discrete distribution trend in the HSV space. The value of

(a) RGB image



(b) Color histogram in CIE L*a*b* color space



(c) Color histogram in HSV color space

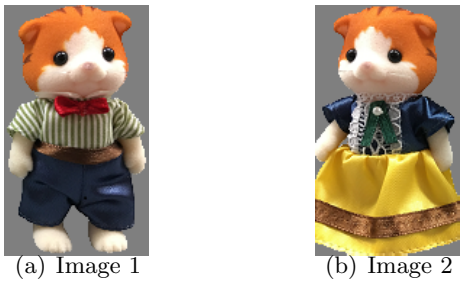
Figure 2.1: Analysis of the distribution of colors in different color spaces.

L* channel defines from 0 and 100. The a* and b* channels mean colors from green

to red and blue to yellow, respectively. The range of these two channels is −128 to 127. CIE L*a*b* color space is closer to human vision. It can separate the lightness independently. To avoid the effect of illumination, the a* and b* channels without the L* channel are used in this study.

2.3 Color histogram decomposition

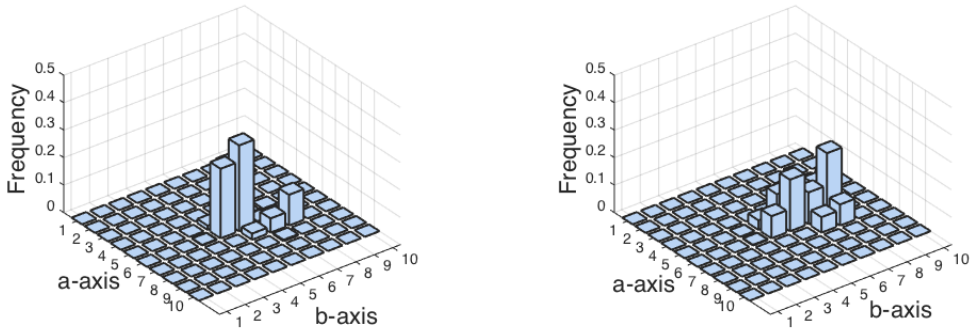
In this section, we explain in detail how to decompose the two-dimensional color histogram and invert absent colors. Fig. 2.2 shows two images 1 and 2 as an example. Let their color space have $I \times J$ bins or quantization. For images 1 and



(a) Image 1 (b) Image 2

Figure 2.2: Images 1 and 2 at 173 × 100 pixels.

2 to be evaluated as matched or unmatched, their two-dimensional color histograms $H = \{h_{ij}\}_{(i,j)=(1,1),\cdots,(I,J)}$ and $G = \{g_{ij}\}_{(i,j)=(1,1),\cdots,(I,J)}$ after normalization are represented in terms of the relative frequencies of classified colors in their bins, confirming their summation to be unity. Moving forward, for explanation purposes, we use H as a representative example. Fig. 2.3 shows the relative histograms H and G for images 1 and



(a) H for image 1 (b) G for image 2

Figure 2.3: Original color histograms ($I = 10, J = 10$).

2, respectively. From the original two-dimensional color histogram H, we create a pair of two-dimensional histograms, $H = {}^{\mathrm{AP}}H + {}^{\mathrm{AB}}H$, ${}^{\mathrm{AP}}H = \{{}^{\mathrm{AP}}h_{ij}\}$, and ${}^{\mathrm{AB}}H = \{{}^{\mathrm{AB}}h_{ij}\}$

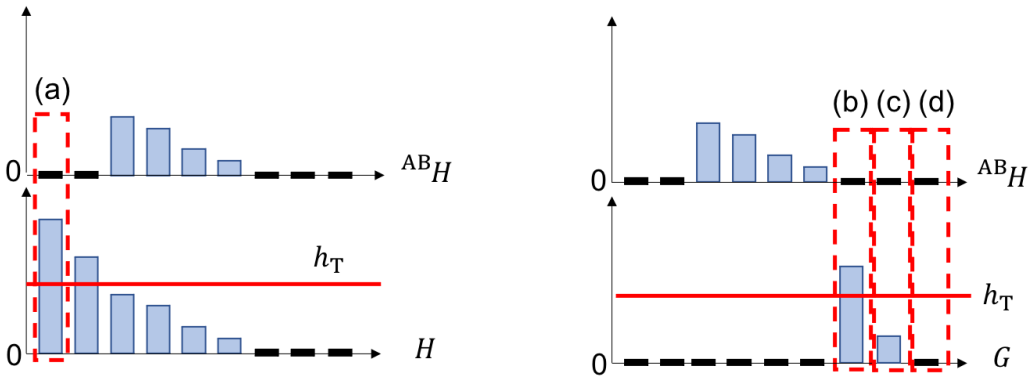
as complements. We have omitted the subscripts i and j in the formula for simplicity and to avoid confusion.

$$^{\mathrm{AP}}h = h(1 - \phi(h)) = h\overline{\phi(h)}, \tag{2.1}$$

$$^{\mathrm{AB}}h = h\phi(h), \tag{2.2}$$

where $\phi(x)$ is an indicator function that shows classification such that $\phi(x) = 1$ if $x \leq h_{\mathrm{T}}$ and $\phi(x) = 0$ otherwise. The threshold value h_{T} is an important parameter in this study and is defined in Section 2.4. Note that the initial values of the other elements without any indication in the above definition were set to 0. $^{\mathrm{AP}}H$ includes major colors frequently observed in image 1, while $^{\mathrm{AB}}H$ contains minor colors that are denoted as "absent" colors because their occurrence is infrequent within the image. Both have the same structure as the two-dimensional histogram H. The elements $^{\mathrm{AP}}h$ and $^{\mathrm{AB}}h$ represent the color frequencies. We expect to systematically and effectively utilize information included in low frequencies in the histogram through the decomposition process.

From $^{\mathrm{AB}}H$, we intend to make an opposite counterpart for the major color histogram as a complementary feature in the original histogram. However, some special cases, such as zero frequency, should be considered. To briefly explain the cases of zero frequency ($^{\mathrm{AB}}h = 0$) during the inversion process of the absent color histogram $^{\mathrm{AB}}H$, we temporarily assume that $^{\mathrm{AB}}H$, H, and G are one-dimensional histograms only in Fig. 2.4. In the



(a) (b)

Figure 2.4: Some special cases of zero frequency $^{\mathrm{AB}}h = 0$ during the inverting process.

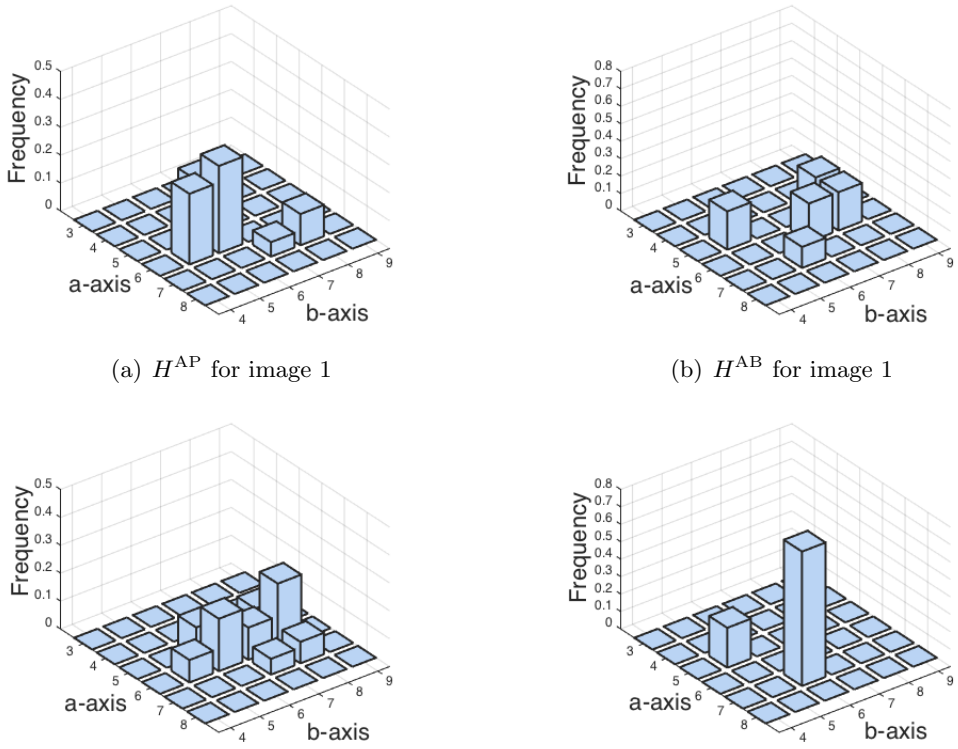
first case (a) shown in Fig. 2.4, if $^{\mathrm{AB}}h = 0$ and $h > h_{\mathrm{T}}$, then $^{\mathrm{AB}}\bar{h} = 0$; in the second case (b), if $^{\mathrm{AB}}h = 0$, $h = 0$ and $g > 0$, then $^{\mathrm{AB}}\bar{h} = h_{\mathrm{T}}$; and, finally, in the third case (c), if $^{\mathrm{AB}}h = 0$, $h = 0$ and $g = 0$, then $^{\mathrm{AB}}\bar{h} = 0$. Table 2.2 summarizes these three transformations to provide a clear explanation. After the abovementioned inverting process, the absent color histogram $^{\mathrm{AB}}\bar{H} = \left\{^{\mathrm{AB}}\bar{h}_{ij}\right\}$ is defined to represent small or zero frequencies in the original one as follows:

$$^{\mathrm{AB}}\bar{h} = (h_{\mathrm{T}} - ^{\mathrm{AB}}h)\phi(h)\psi(h) + h_{\mathrm{T}}\overline{\psi(h)}\psi(g). \tag{2.3}$$

Table 2.2: Value of ${}^{\mathrm{AB}}\bar{h}$ in conditions of ${}^{\mathrm{AB}}h = 0$.

${}^{\mathrm{AB}}\bar{h}$	${}^{\mathrm{AB}}h = 0, h > h_{\mathrm{T}}$	${}^{\mathrm{AB}}h = 0, h = 0$
$g > 0$	0	h_{T}
$g = 0$	0	0

Here, $\psi(x)$ is another indicator function that satisfies the following conditions: $\psi(x) = 1$ if $x > 0$; otherwise, $\psi(x) = 0$. Finally, it is necessary to normalize both ${}^{\mathrm{AP}}H$ and ${}^{\mathrm{AB}}\bar{H}$ into H^{AP} and H^{AB} to satisfy the condition that all components should sum to 1. Fig. 2.5



(a) H^{AP} for image 1

(b) H^{AB} for image 1

(c) G^{AP} for image 2

(d) G^{AB} for image 2

Figure 2.5: Apparent and absent color histograms.

shows apparent or major color histograms H^{AP} and G^{AP} and absent color histograms H^{AB} and G^{AB} for images 1 and 2, respectively.

2.4 Threshold definition

In the algorithm described in the last section, threshold h_{T} plays a main role in defining apparent and absent colors [77]. This section describes how to define h_{T} to provide a meaningful algorithm with effective performance. We first introduce the *mean color histogram* M to find an averaged tendency of the color distributions in two histograms to be compared, and then to realize a stable definition of the threshold. $M = \{m_{ij}\}_{(i,j)=(1,1),\cdots,(I,J)}$ is defined as

$$m_{ij} = \frac{h_{ij} + g_{ij}}{2}. \tag{2.4}$$

Generating mean color histogram is a critical phase before threshold selection. The proportion of each color in the histogram is statistically analyzed for matching images, thereby improving the rationality and dynamism of threshold determination and guaranteeing accuracy of the final similarity measurement. We next convert the two-dimensional histogram M to a sorted one-dimensional histogram M^{sorted} as

$$M^{\mathrm{sorted}} = \left\{ m_{i-1}^{\mathrm{sorted}} \geq m_i^{\mathrm{sorted}} \right\}. \tag{2.5}$$

The threshold value h_{T} can be defined by the following equation through use of an order index s related to a significant rate α, by which we can separate the set of all bins into sets of apparent and absent colors by considering the rarity of absent colors in the images.

$$h_{\mathrm{T}} = \frac{m_s^{\mathrm{sorted}} + m_{s+1}^{\mathrm{sorted}}}{2}, \tag{2.6}$$

$$s = argmin \left\{ \sum_{i=1}^{s} m_i^{\mathrm{sorted}} \geq 1 - \alpha \right\}. \tag{2.7}$$

By using parameter s, a stable decomposition can be performed with no "chattering" near the threshold value in comparison with a constant threshold. Fig. 2.6 shows the mean color histogram for histograms H and G. Fig. 2.7 is a Pareto chart [78] for this example. We can use this information to determine the significant rate α, which represents the effectiveness at revealing the rareness of absent colors and contributes to setting of the threshold value.

Fig. 2.7 shows the distribution of M^{sorted} when the distributions of two color histograms are similar. In contrast, Fig. 2.8 shows the case of opposite distributions, where the distribution of M^{sorted} shows a uniform distribution in the histogram. We can observe that the frequency in all bins has a smaller change. If the smallest frequency in the histogram exceeds α, then no bins are assigned to absent colors, so no absent colors need to be added to the calculation for the similarity measure. When the smallest frequency is less than or equal to α, the inverse of the H and G distributions can still reduce similarity of the results, because the inverted low frequencies are in different bins in the two original color histograms. Note that the mean color histograms can be used both to derive the threshold h_{T} and to calculate similarities based on apparent and absent

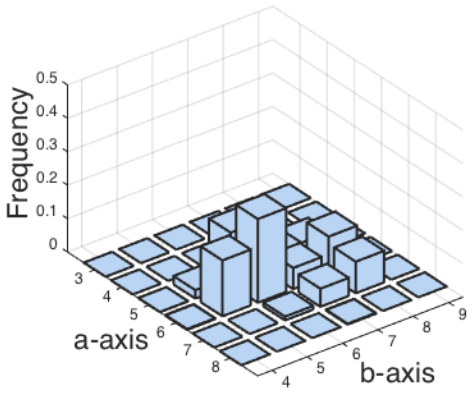


Figure 2.6: Mean color histogram.

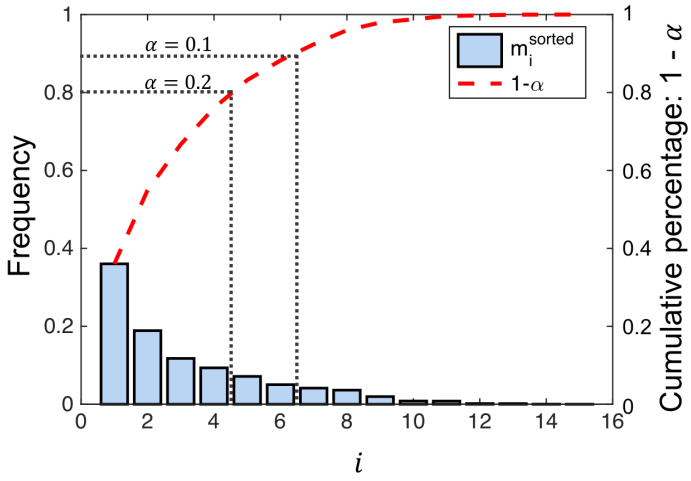


Figure 2.7: Pareto chart of parameter α and sorted histogram M^{sorted}.

colors, as described in the next section. In this work, we proposed a novel approach using a method called ABC for robust pattern search or matching. A key contribution of this work is it offers a fair way of focusing on contributions of absent colors belonging to low-frequency or vacant bins in the histograms. Furthermore, we also introduce a mean color histogram into the ABC to find an averaged tendency of the color distributions in two histograms to be compared, and thereafter to realize a stable definition of the threshold. Finally, we evaluate a variety of similarity measures that can be combined with the ABC method. In Section 2.6, the experimental results of Mondrian random patterns, real-world images, and open data demonstrate the effectiveness of our method for matching even in some difficult cases.

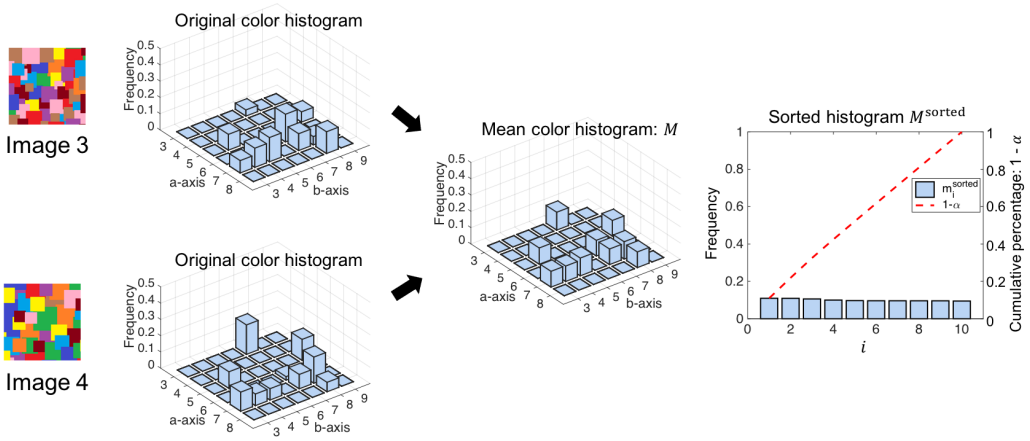


Figure 2.8: Analysis of the case of uniform distribution in M.

2.5 Similarity measure analysis of ABC

Many measures for testing similarity between two histograms or probability density functions have been proposed, including intersection, chi-square distance, Jensen–Shannon divergence, and Bhattacharyya distance. This section describes these measures in combination with our ABC to show its universality in evaluating the similarity of images as described in Section 2.6.1. We expect this universality of ABC to make it useful as an effective schema for many applications.

2.5.1 Intersection

Intersection [79, 80, 81] has been used in many studies and applications because of its simplicity. It is defined for two same-sized histograms H and G as

$$I(H,G) = \sum_{(i,j)=(1,1)}^{(I,J)} min\{h_{ij}, g_{ij}\}. \tag{2.8}$$

For the two histogram types proposed in this paper, we define a scheme for combining the two intersections by use of weighting coefficients as [82]

$$S = w_{\mathrm{AP}} I\left(H^{\mathrm{AP}}, G^{\mathrm{AP}}\right) + w_{\mathrm{AB}} I\left(H^{\mathrm{AB}}, G^{\mathrm{AB}}\right), \tag{2.9}$$

where w_{AP} and w_{AB} are weights for balancing to the two types of intersections by the constraint $w_{\mathrm{AP}} + w_{\mathrm{AB}} = 1$.

2.5.2 Chi-square distance

The chi-square test is a non-parametric test mainly used to compare two or more sample rates. To measure similarity between two histograms, we use the χ^2 statistic to observe frequencies. We define our version using the mean color histogram as

$$\chi^2(H,G) = \sum_{(i,j)=(1,1)}^{(I,J)} \frac{(h_{ij} - m_{ij})^2}{m_{ij}} + \sum_{(i,j)=(1,1)}^{(I,J)} \frac{(g_{ij} - m_{ij})^2}{m_{ij}}, \tag{2.10}$$

where m_{ij} is the element of the mean color histogram. Eq. (2.10) can be simplified as

$$\chi^2(H,G) = \sum_{(i,j)=(1,1)}^{(I,J)} \frac{(h_{ij} - g_{ij})^2}{(h_{ij} + g_{ij})}. \tag{2.11}$$

The total distance between the histograms is defined as follows by using the weights:

$$D_{\chi^2} = w_{\mathrm{AP}}\chi^2\left(H^{\mathrm{AP}}, G^{\mathrm{AP}}\right) + w_{\mathrm{AB}}\chi^2\left(H^{\mathrm{AB}}, G^{\mathrm{AB}}\right). \tag{2.12}$$

2.5.3 Jensen–Shannon divergence

JS divergence [83] is a symmetric divergence measurement based on Kullback–Leibler divergence [84]. By calculating divergences between histograms, a larger divergence indicates smaller correlative relation and smaller similarity between the histograms. This is defined as

$$D_{\mathrm{KL}}(H \parallel G) = \sum_{(i,j)=(1,1)}^{(I,J)} h_{ij} log \frac{h_{ij}}{g_{ij}}. \tag{2.13}$$

We define a particular version of JS divergence as follows by using the mean color histogram M:

$$D_{\mathrm{JS}}(H,G) = \frac{1}{2}D_{\mathrm{KL}}(H \parallel M) + \frac{1}{2}D_{\mathrm{KL}}(G \parallel M). \tag{2.14}$$

In this formula, the antilogarithm cannot be zero in logarithm calculations so that we can give $log\frac{h_{ij}}{g_{ij}}$ a relatively minimum value. The JS divergence-based distance between the histograms is defined as

$$D_{\mathrm{JSD}} = w_{\mathrm{AP}}D_{\mathrm{JS}}\left(H^{\mathrm{AP}}, G^{\mathrm{AP}}\right) + w_{\mathrm{AB}}D_{\mathrm{JS}}\left(H^{\mathrm{AB}}, G^{\mathrm{AB}}\right). \tag{2.15}$$

2.5.4 Bhattacharyya distance

In statistics, the Bhattacharyya distance [85] is often used to measure dissimilarity of two discrete or continuous probability distributions. It is closely related to the Bhattacharyya coefficient, which measures overlap between two statistical samples or populations. Bhattacharyya distance can also be used to determine relative relationships between two samples or to determine differences between two classes. We can consider the two histograms as discrete probability distributions, so the formula is

$$BD = -ln(BC(H,G)), \tag{2.16}$$

where $0 \leq BD \leq \infty$, $0 \leq BC \leq 1$, and $BC(H, G)$ is the Bhattacharyya coefficient,

$$BC(H, G) = \sum_{(i,j)=(1,1)}^{(I,J)} \sqrt{h_{ij} g_{ij}}. \tag{2.17}$$

The distance between two sets of apparent and absent color histograms is defined as

$$D_{\mathrm{BD}} = w_{\mathrm{AP}} BD\left(H^{\mathrm{AP}}, G^{\mathrm{AP}}\right) + w_{\mathrm{AB}} BD\left(H^{\mathrm{AB}}, G^{\mathrm{AB}}\right). \tag{2.18}$$

Algorithm 1: Proposed ABC approach

Input: Reference image S_{R} and compared image S_{S}.

Output: Target location L_{T} in the scene.

Initialization: $I = 10, J = 10, \alpha = 0.2, w_{\mathrm{AP}} = 0.6, w_{\mathrm{AB}} = 0.4$

for *each* $S_{\mathrm{S}}(i, j)$ **do**

- Crop the compared image from position (i, j) in the scene.
- Generate two-dimensional color histograms H and G by a* and b* channels.
- Divide color histograms into apparent color histograms ${}^{\mathrm{AP}}H$, ${}^{\mathrm{AP}}G$ and absent color histograms ${}^{\mathrm{AB}}H$, ${}^{\mathrm{AB}}G$.
- Invert absent color histograms ${}^{\mathrm{AB}}H$ and ${}^{\mathrm{AB}}G$ to ${}^{\mathrm{AB}}\bar{H}$, ${}^{\mathrm{AB}}\bar{G}$.
- Normalize apparent and absent color histograms $\left\{H^{\mathrm{AP}}, G^{\mathrm{AP}}, H^{\mathrm{AB}}, G^{\mathrm{AB}}\right\}$.
- Calculate similarity $R_{(i,j)}$ by H^{AP} and G^{AP}, H^{AB} and G^{AB}.

All locations are scanned, then find position L_{T} with $max(R_{(i,j)})$;

2.6 Ability of ABC

We first describe a color image $I_{\mathrm{t}} = \{t_{ij}\}$. We denote I_{t} by signal $I_{\mathrm{s}} = \{s_{ij}\}$ and noise $I_{\mathrm{n}} = \{n_{ij}\}$, is defined as

$$I_{\mathrm{t}} = I_{\mathrm{s}} + I_{\mathrm{n}}. \tag{2.19}$$

The variance of image I_{t} is represented as

$$\sigma_{\mathrm{t}}^2 = E\left(I_{\mathrm{t}} - \mu_{\mathrm{t}}\right)^2 = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \left[I_{\mathrm{t}}(i, j) - \mu_{\mathrm{t}}\right]^2, \tag{2.20}$$

where $\mu_{\mathrm{t}} = \mu_{\mathrm{s}} + \mu_{\mathrm{n}}$ is the average value in image I_{t}. μ_{s} and μ_{n} are the average values of I_{s} and I_{n}, and $E(I_{\mathrm{n}}) = \mu_{\mathrm{n}} = 0$ describes the noise distributions obeyed by some balanced and unbiased Gaussian distributions. m and n are the sizes of the images. As

Table 2.3: Statistical characteristics of background noise

		R	G	B
Biker	σ_{n}	0.958	0.903	1.143
	SNR	35.09	35.20	34.17
Walking	σ_{n}	1.021	0.953	1.028
	SNR	36.48	37.90	37.23

mentioned in Eq. 2.20,

$$\begin{aligned}\sigma_{\mathrm{t}}^2 &= E\left((I_{\mathrm{s}}+I_{\mathrm{n}})-(\mu_{\mathrm{s}}+\mu_{\mathrm{n}})\right)^2 \\ &= E\left((I_{\mathrm{s}}-\mu_{\mathrm{s}})^2\right)+2E\left((I_{\mathrm{s}}-\mu_{\mathrm{s}})(I_{\mathrm{n}}-\mu_{\mathrm{n}})\right)+E\left((I_{\mathrm{n}}-\mu_{\mathrm{n}})^2\right) \\ &= E\left((I_{\mathrm{s}}-\mu_{\mathrm{s}})^2\right)+E\left((I_{\mathrm{n}}-\mu_{\mathrm{n}})^2\right) \\ &= \sigma_{\mathrm{s}}^2+\sigma_{\mathrm{n}}^2,\end{aligned} \tag{2.21}$$

where σ_{s}^2 and σ_{n}^2 represent variance of the signal and noise in image I_{s} and I_{n}. The formulas are calculated as

$$\sigma_{\mathrm{s}}^2=\frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left[I_{\mathrm{s}}(i,j)-\mu_{\mathrm{s}}\right]^2, \tag{2.22}$$

$$\sigma_{\mathrm{n}}^2=\frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left[I_{\mathrm{n}}(i,j)-\mu_{\mathrm{n}}\right]^2. \tag{2.23}$$

The signal-to-noise ratio SNR is calculated as

$$SNR=10\log\frac{\sigma_{\mathrm{s}}^2}{\sigma_{\mathrm{n}}^2}=10\log\frac{\sigma_{\mathrm{t}}^2-\sigma_{\mathrm{n}}^2}{\sigma_{\mathrm{n}}^2}=10\log\left(\frac{\sigma_{\mathrm{t}}^2}{\sigma_{\mathrm{n}}^2}-1\right). \tag{2.24}$$

Next, we analyzed standard deviations σ_{n} of noise in stationary background regions of the datasets [86], where we collected the small regions of interest (ROI) with sizes of $(100, 100)$ at the same positions in their background without any moving things: $(61, 531)$ in 50 frames of *Biker*, and $(1, 1)$ in 50 frames of *Walking*, respectively. σ_{n} can be calculated from numerous estimated noise values in the difference images $I_{\mathrm{n}}=I_{\mathrm{t}}-\bar{I}_{\mathrm{t}}$. σ_{t} is from one of 50 images in each dataset.

Table 2.3 shows the SNR in the sampled images of these databases. The average value of SNR in these two datasets is 36.01. To add noises to the noise-free Mondrian random pattern, we utilize the knowledge on noise obtained from Table 2.3 and thereafter add the noise of unbiased Gaussian distributions to the individual channels of the color image by $SNR = 36$, 33, and 30 where a reduced SNR implies increased noise. The example model of how to add noise into the noise-free color image is shown in Fig. 2.9.

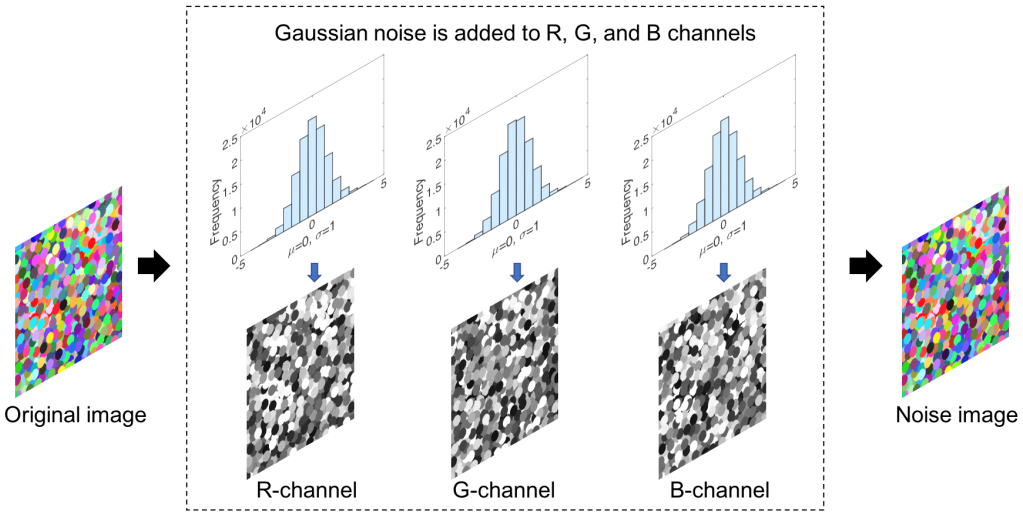


Figure 2.9: The example model of adding noise.

2.6.1 Basic analysis on Mondrian pattern

In this section, for performance evaluation we compare ABC with three other color histogram-based matching methods, CI [8], CCH [9], and TFCM [52] because these methods are frequently cited and used in many papers and studies. The reference image and scene are shown in Fig. 2.10. We next add noise to the color image. We first separate the individual channels of the color image and then add unbiased Gaussian noise of SNR = 36, 33, and 30 to each channel, before finally merging the channels into a color image. Therefore, CI, CCH, TFCM, and ABC are tested by use of these Mondrian random patterns embedded as typical additional noise.

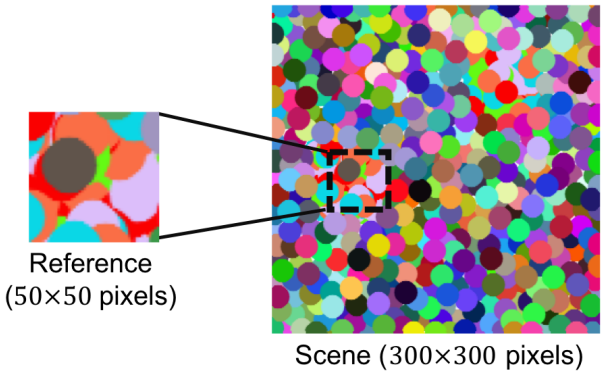


Figure 2.10: Mondrian random pattern (noise-free).

Margin [87] is defined as the difference in similarity between the best and second-best matching positions. The best-matching position is the maximum value at which the area

of IoU between ground truth and the searched position should exceed 0.90. The second-best matching position is the maximum value at which the area of IoU is less than 0.15. A larger margin provides stronger distinguishability for avoiding interference by similar objects. Fig. 2.11 shows example profiles. They are projected onto the horizontal axis by 3D similarity profiles for ABC, CI, CCH, and TFCM. We found that the margin of ABC was much larger than that of the other three methods. We compared the best-

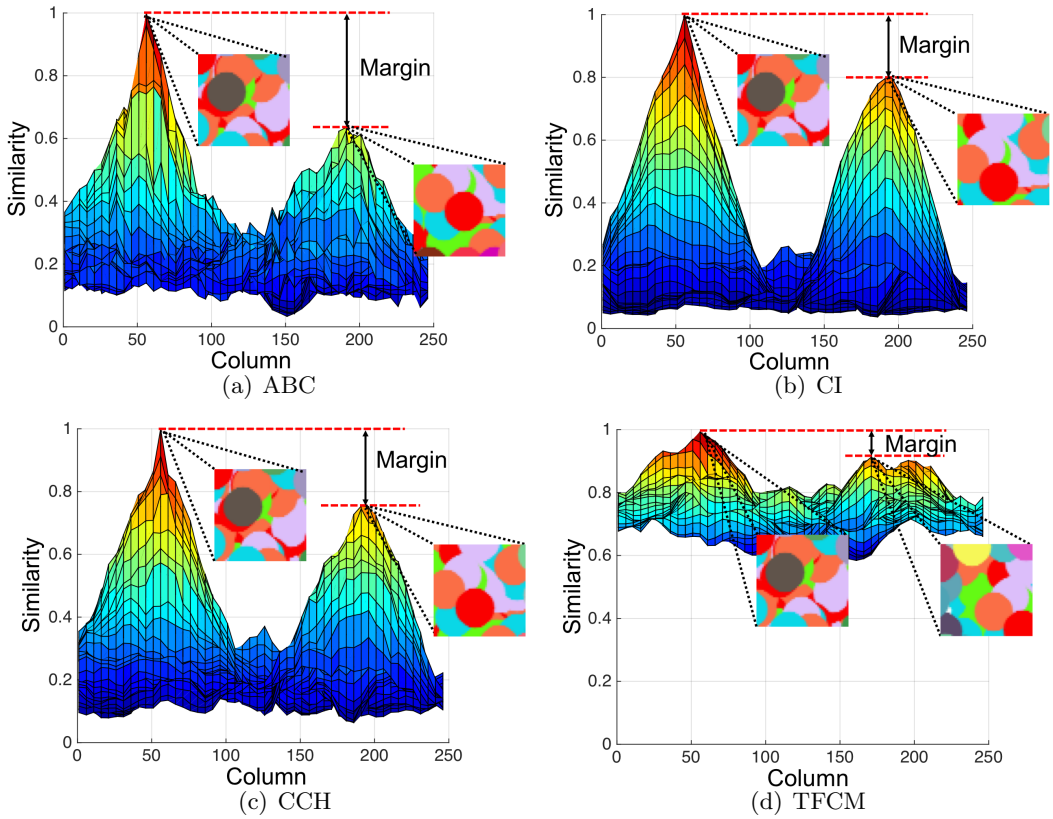


Figure 2.11: Projected profiles of similarity (No noise $\sigma = 0$).

matched image at the highest peak and the second-best matched image, and observe the apparent and absent colors by their visibility maps, as shown in Fig. 2.12. Fig. 2.12 includes two matching examples, the one at the second-best position and the other at randomly selected position for showing details of the histogram decomposition, where we have three original images, i.e. the reference image at the position (128, 55), the second-best image, and then the random image all in Fig. 2.10. Here, we do not include the best-matched image because we use the target scene of no noise so it must be the same as the reference. The second column shows apparent color visibility maps including $\{a, b, c\}$, where the white color indicates absent or non-existing colors, while the fourth and fifth columns show absent color visibility maps with respect to two conditions, the one defined by low-frequencies as $\{p1, q1, r1, s1\}$ and the other by non-existing colors as $\{p2, p3, q2, q3, r2, r3, s2, s3\}$. The third and sixth columns are their histograms,

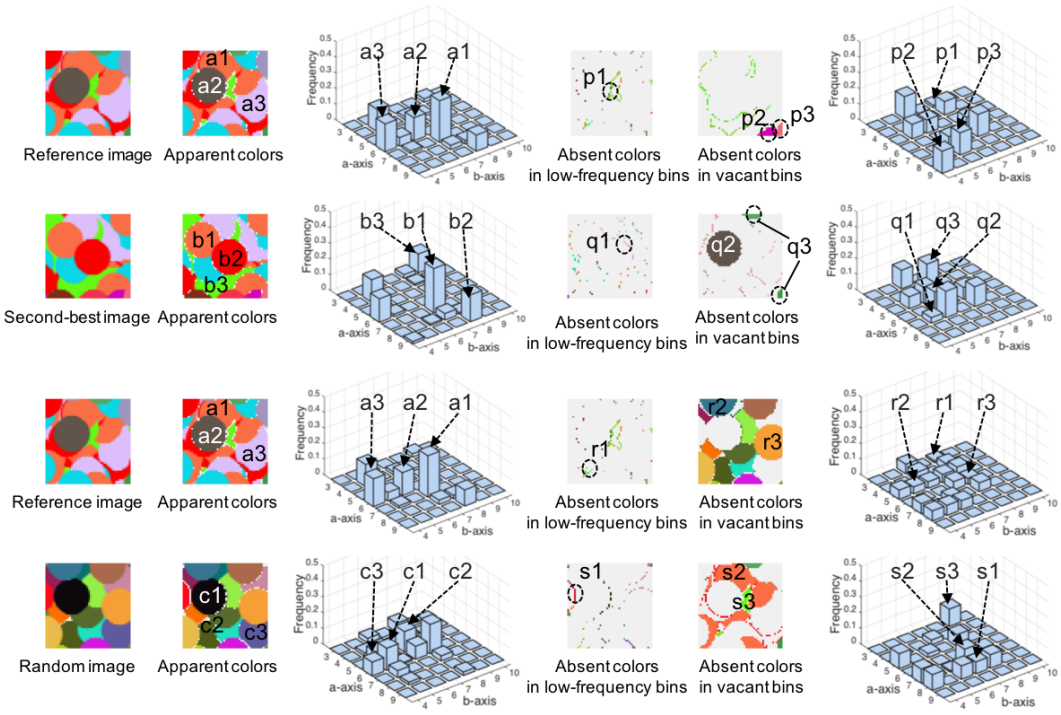


Figure 2.12: Visibility map for apparent and absent colors.

respectively. We can observe in the first row the apparent colors $a1$, $a2$ and $a3$ are in the position (7, 7), (6, 6) and (6, 4), and the absent colors $p1$, $p2$ and $p3$ are in the position (3, 8), (9, 4) and (8, 6) of their own histograms, respectively. We can evaluate that the absent colors $p2$, $p3$, $q2$, and $q3$ are of a lot importance in the contribution to high similarity margins by use of the proposed method. In the last two rows at the bottom, the combination of the reference and a randomly selected image makes a lot of absent colors in their histograms and images representing the colors that are not included in the original images at all, which do not have any large overlap in their histograms. Observing the absent color maps, we also see that some shapes in connected parts have some absent colors, mainly because of artifactual color mixture caused by smoothing in the process of color conversion. Absent colors thus capture this subtle change with high accuracy, providing a more reliable analysis of color information in the image. This case could show possible examples in which absent colors contribute to decreasing similarity, resulting in a larger margin through their use.

To statistically investigate the performance of ABC and others in terms of margin, we independently extracted from random positions 100 reference images sized 50×50, and then we searched for them within the reference image itself or in contaminated versions with noise. For each, we could find the best position or peak as $peak1$ and the second-best position as $peak2$. From the noisy image, we separated 100 pairs of $peak1$ and $peak2$, and thereafter created their histograms as shown in Fig. 2.13. The similarity

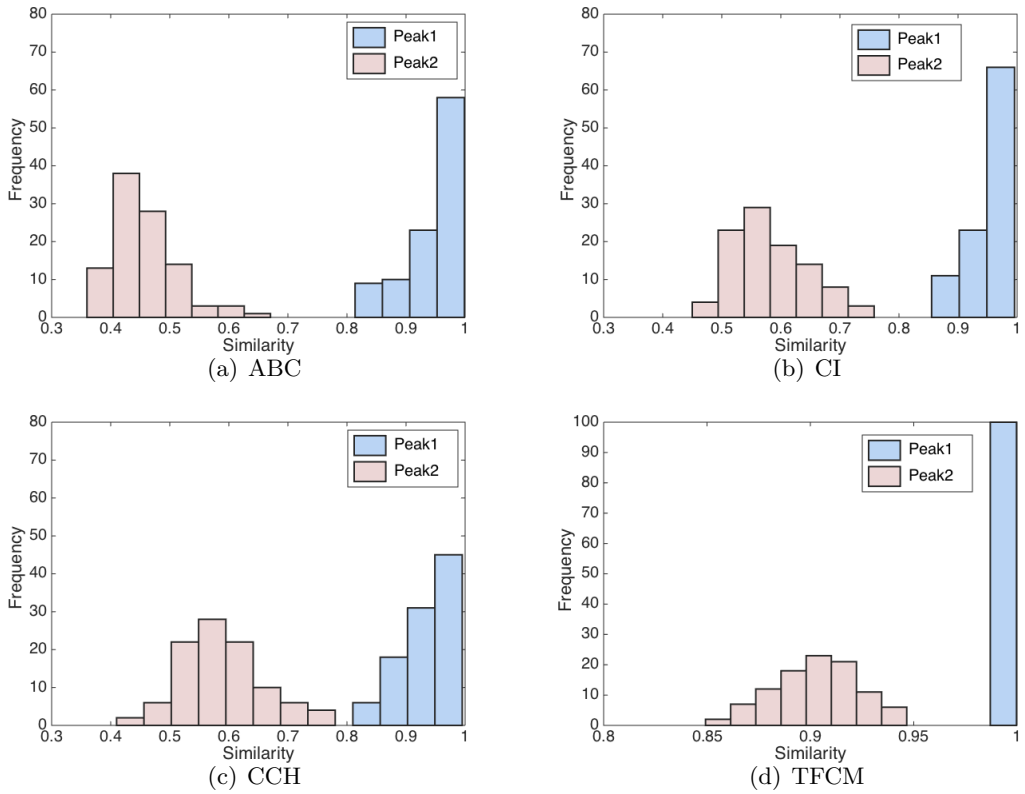


Figure 2.13: Distributions of margins between *peak*1 and *peak*2 for ABC, CI, CCH, and TFCM (*SNR*=36).

ranges of ABC, CI, and CCH are the same from 0.3 to 1, and that of TFCM is from 0.8 to 1. To better clarify this observation, we utilize Fisher's ratio (*FR*) to evaluate the discriminating power of any two-class discriminator. *FR* is defined as

$$FR = \frac{(m_{\mathrm{p1}} - m_{\mathrm{p2}})^2}{\sigma_{\mathrm{p1}}^2 + \sigma_{\mathrm{p2}}^2}, \tag{2.25}$$

where m_{p1}, m_{p2}, σ_{p1}^2, and σ_{p2}^2 represent the means and variances, respectively, of the *peak1* and *peak2* classes. Here, an increased numerator indicates an increased distance or interval between classes, and a reduced denominator indicates the greater compactness of each class; thus, *FR* can be greater when the feature space for classification is better for discrimination or identification. Table 2.4 summarizes *FR* obtained by CI, CCH, TFCM, and ABC in cases of noisy Mondrian random patterns, where we vary the SNR is from 36 to 30. Table 2.4 indicates that *FR* of ABC was superior to those of CI, CCH, and TFCM with the case of $SNR = 36$, 33, and 30, and furthermore, that it maintained the performance even in the case of a reduced *SNR*. As shown in Fig. 2.14, the horizontal axis represents the change in variance, and the vertical axis represents the change in mean value, which is the size of the average margin. In these experiments, we use every set of ten images for observing some distributions in the means and variances that construct

Table 2.4: Margins in CI, CCH, TFCM, and ABC

		CI		CCH		TFCM		ABC	
SNR	Class	Mean	Std	Mean	Std	Mean	Std	Mean	Std
36	Peak1	0.952	0.034	0.935	0.041	0.999	3.6×10^{-4}	0.946	0.049
	Peak2	0.583	0.060	0.590	0.069	0.902	0.020	0.456	0.055
	FR	28.67		18.28		21.75		**43.77**	
33	Peak1	0.937	0.038	0.915	0.042	0.999	2.2×10^{-4}	0.935	0.057
	Peak2	0.575	0.058	0.591	0.072	0.910	0.021	0.455	0.064
	FR	26.52		14.80		16.95		**31.18**	
30	Peak1	0.919	0.040	0.898	0.045	0.998	5.7×10^{-4}	0.914	0.058
	Peak2	0.579	0.053	0.592	0.065	0.908	0.024	0.465	0.064
	FR	25.79		14.85		14.07		**26.58**	

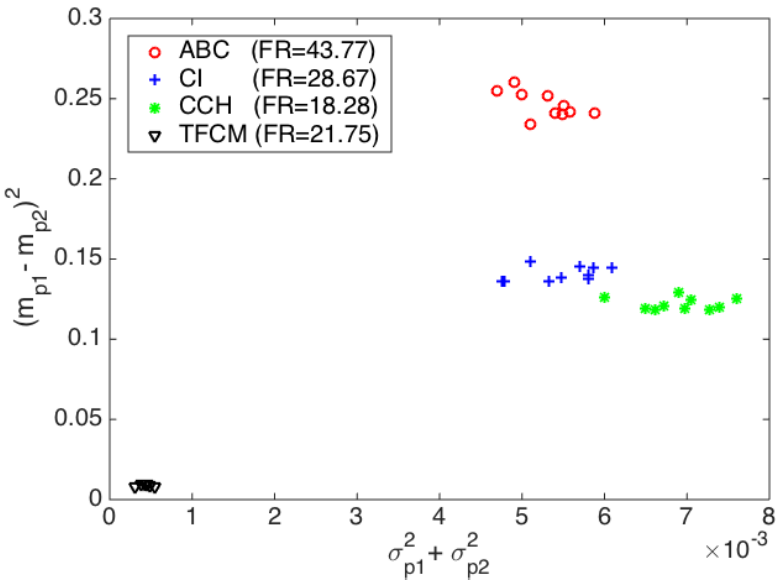


Figure 2.14: Analytical diagram of FR (SNR=36).

their FR values. As a result, we may have two types of similarity measures corresponding to the two approaches to increase FR: to increase the difference or distance between the means or centers of the two classes and to decrease their own variances. ABC adopts the former approach, while TFCM uses the latter approach.

2.6.2 Verification of ABC's performance

Search in a cluttered scene

We designed search tasks in a cluttered scene and with actual objects under five conditions: varied illumination, rotation, deformation, scaling, and occlusion. In these

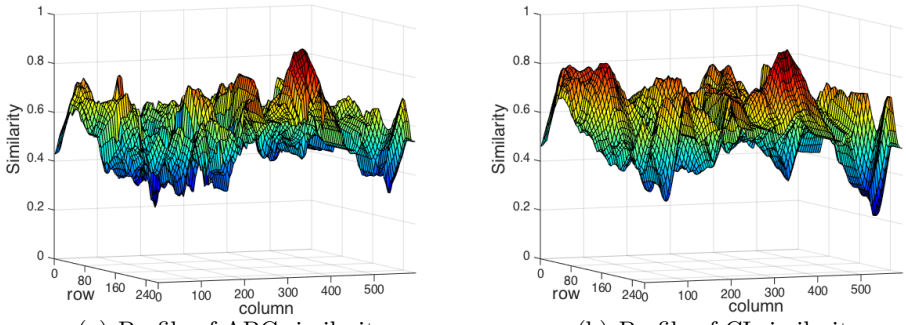
(a) Reference (b) The cluttered scene under dark illumination

Figure 2.15: Search in a cluttered scene. The red, blue, and green bounding boxes respectively show ABC, CI, CCH, and TFCM.

experiments, the reference in Fig. 2.15(a) is of size 88×44, and the five scenes are of the same size, 360×640. We first tested ABC performance under the illumination change as shown in Fig. 2.15. Except for TFCM, the other three methods could correctly find the position of the reference under the illumination change, but observing their profiles, we can see that the difference between the highest and second-highest peak under ABC-based search was larger than that under the other two. This observation can be confirmed from its profile in Fig. 2.16(a), which also verifies that ABC may have the best discrimination performance among these classes. We compare the highest peak and the second highest peak of the profile graphs, ABC has a larger margin. Fig. 2.17 shows the other challenges of searching for the reference. In Fig. 2.17(a), the doll was rotated by a small angle at the same location. In Fig. 2.17(b), the doll's clothing was deformed, and in Fig. 2.17(c) the doll ' s apparent size was reduced by moving it back. Finally, in Fig. 2.17(d) the doll was occluded by placing it behind another doll, but ABC and others could still find its true position.

2.6.3 Tracking in open data

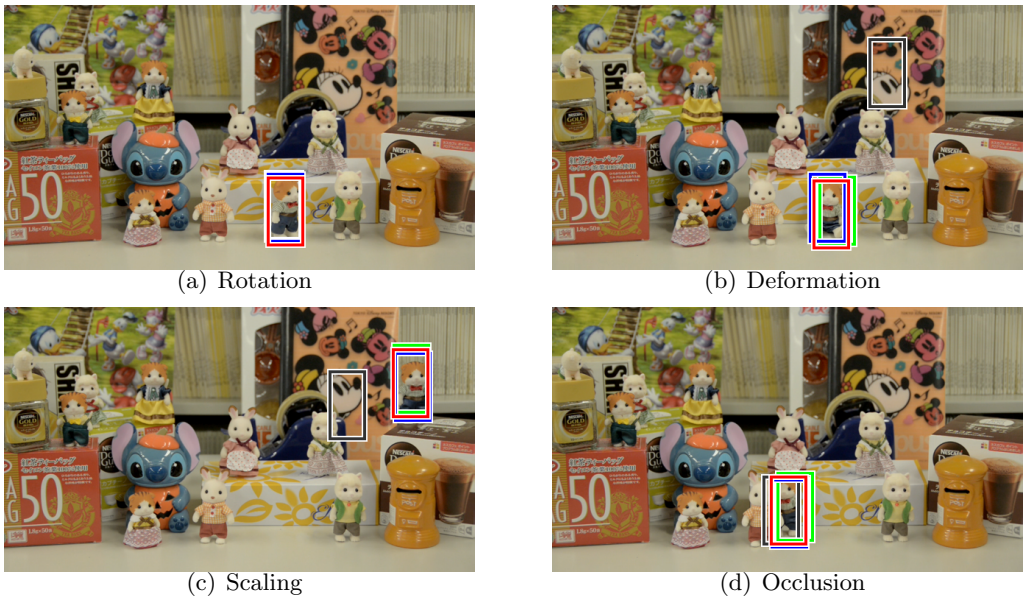
Tracking is a difficult task requiring a stable and reliable matching scheme for consecutive following of objects of interest. Many conventional tracking algorithms [88, 89, 90] track targets by accurately updating the reference, but here our aim was to test the fundamental feasibility of ABC in tracking tasks with no modification of the single reference. *Liquor* [86] is an open dataset that includes interesting sequences of bottles being picked up and moved by human hands, where rotation, deformation, scaling, and occlusion happen over many consecutive frames. We used frames #1301–#1400 to test the tracking



(a) Profile of ABC similarity (b) Profile of CI similarity

(c) Profile of CCH similarity (d) Profile of TFCM similarity

Figure 2.16: Three-dimensional profiles of similarity under the case of illumination change by ABC, CI, CCH, and TFCM.



(a) Rotation (b) Deformation

(c) Scaling (d) Occlusion

Figure 2.17: Search under rotation, deformation, scaling, and occlusion by ABC, CI, CCH, and TFCM.

performances of CI, CCH, TFCM, and ABC. In frames #1301–#1380, the target bottle remains in the leftmost position. Other bottles are picked up by hands, passed in the front of the target bottle, then moved to other positions. In frames from #1381 to the end, the target is picked up and moved forward, including small out-of-plane rotations to the left and right.

In the first frame, we defined the reference shown at the top left in Fig. 2.18. We used

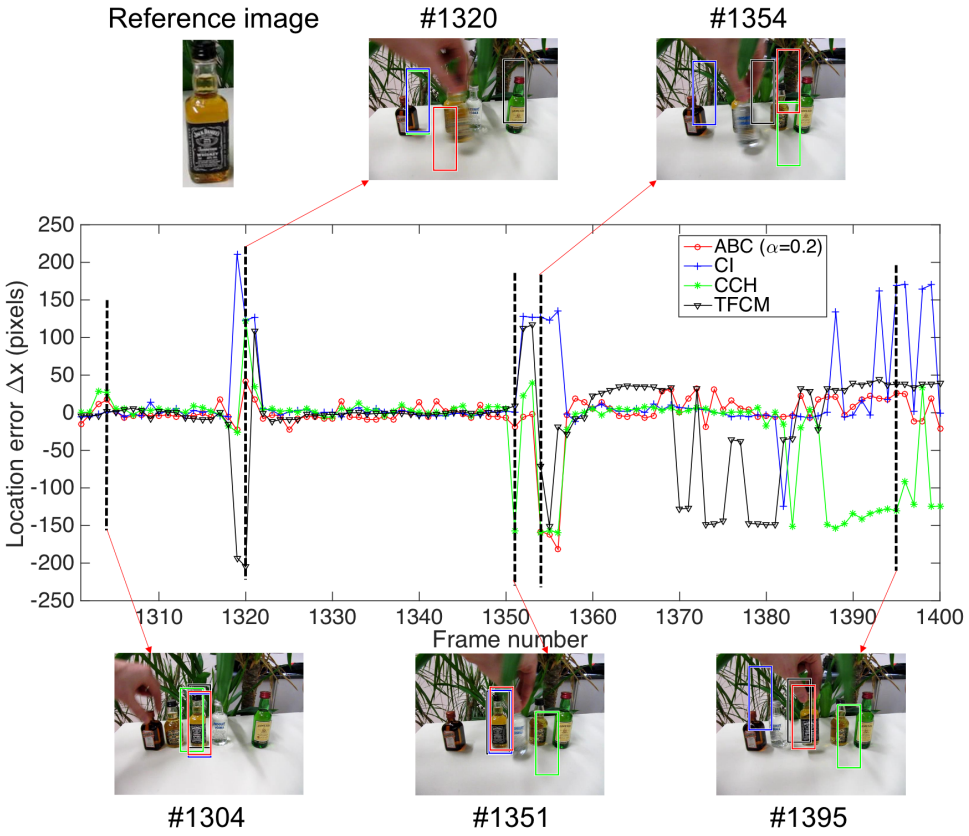


Figure 2.18: Tracking by ABC ($\alpha = 0.2$), CI, CCH, and TFCM.

horizontal displacement Δx of matched positions from the central ground truth positions as a fundamental evaluator, because in these frames the bottle of interest was moved mainly horizontally, and other candidates mismatched bottles were placed on the same level. Therefore, smaller Δx can show better tracking performance in the algorithms. In this tracking experiment, we chose the following five typical situations for the target bottle to demonstrate ABC performance: 1) In frame #1304, all three methods can find reasonable positions, because the bottle is not moved. 2) Because of the large occlusion in frame #1320, only ABC could capture the lower-left position of the target, while other three methods were misled to unreasonable places. 3) Because of the partial occlusion in frame #1351, the result from CCH is largely shifted. 4) In frame #1354, possibly the most difficult case of full occlusion, none of the four methods can identify the target. 5)

In frame #1395, where the bottle is lifted and moved to the right, only ABC maintained stable tracking of the bottle.

Table 2.5: Number of mismatches in 100 frames

	CI	CCH	TFCM	ABC
$\|\Delta x\| \geq 20$	**15**	28	49	16
$\|\Delta x\| \geq 30$	15	22	40	**7**
$\|\Delta x\| \geq 40$	15	20	18	**4**
$\|\Delta x\| \geq 50$	15	19	16	**3**

Table 2.5 shows the number of mismatches in the *Liquor* dataset for frames #1301–#1400. The size of the reference image is 210×73 pixels. If $|\Delta x|$ exceeds 73, the searched position does not overlap with ground truth. We thus set 20, 30, 40, or 50 pixels to observe the number of mismatches. For example, if $|\Delta x|$ exceeds 20, the result is incorrect matching. As a result, ABC could not find targets under large occlusion conditions, but it remains robust in other frames.

In Fig. 2.19, we show the performance of the *Skiing* [86] dataset for frames #31–#60. The reference image is defined at the top left, and the size is 41×39 pixels. The challenges are deformation and rotation. 1) In frame #32, ABC, CI, CCH, and TFCM can search for the correct positions because the skier does not change significantly in the sky. 2) Frames #38, #45, and #58 have only ABC that can match the target. This is because ABC provides a good balance between the major and absent colors. In the comparison process, red and yellow colors are major colors for the reference image, and they are also the absent colors for the compared images. 3) In frame #50, the proportion of yellow pixels is decreased because of the large deformation. Therefore, four methods cannot match the target. Table 2.6 shows the number of mismatches in the *Skiing*

Table 2.6: Number of mismatches in 30 frames

	CI	CCH	TFCM	ABC
$\|\Delta x\| \geq 20$	25	24	23	**5**
$\|\Delta x\| \geq 30$	24	24	23	**4**
$\|\Delta x\| \geq 40$	23	23	23	**4**
$\|\Delta x\| \geq 50$	23	23	23	**4**

dataset for frames #31–#60. The number of mismatches also proves the performance of our proposed ABC approach. ABC is more robust in the matching process.

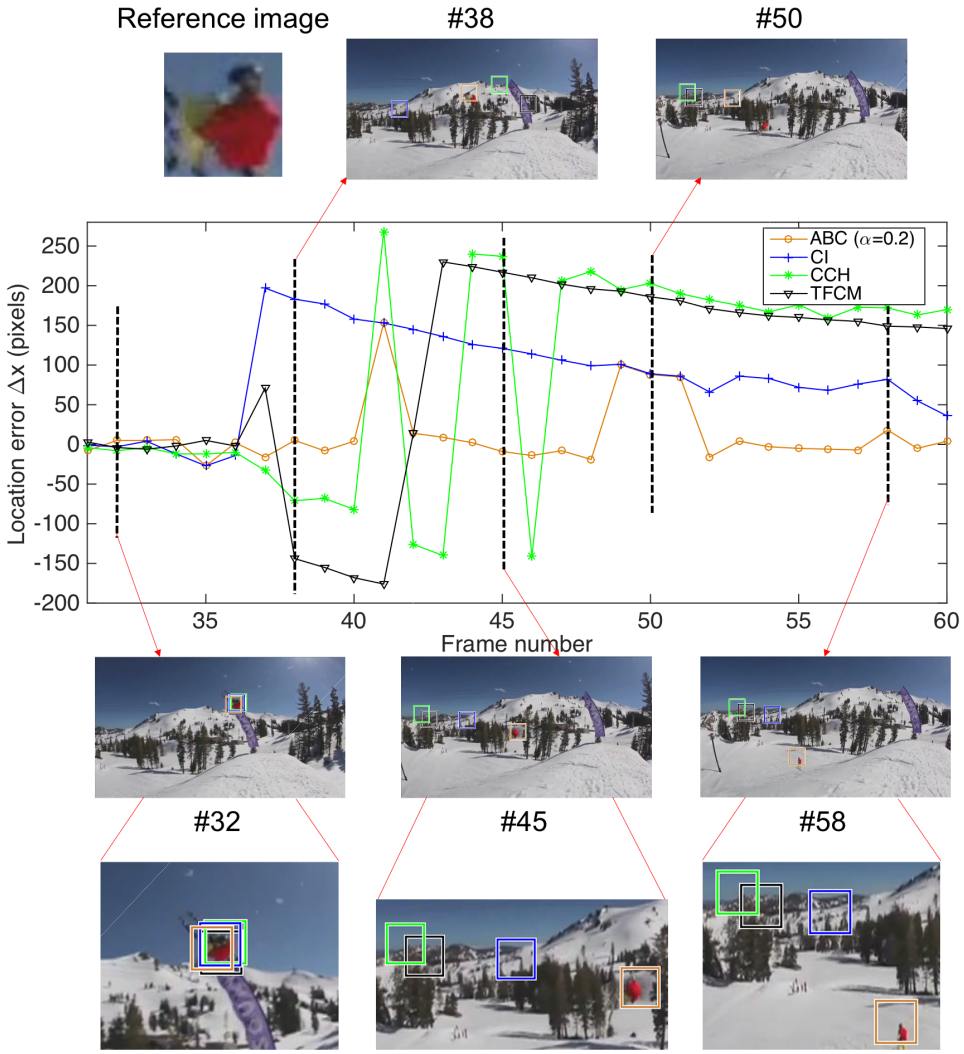


Figure 2.19: Matching performance in *Skiing* dataset.

2.7 Summary

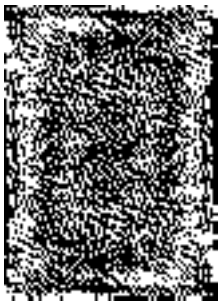
In this chapter, we first introduce the design of our proposed ABC approach. To divide the color histogram into apparent and absent color histograms, we provide a way to obtain threshold h_T by using the mean color histogram. Inverting is an important step for enhancing the performance of absent colors. Thereafter, absent or minor colors are used to provide fair treatment. Finally, four similarity measures were put forward to calculate the similarity of the ABC method. The performance of our proposed ABC approach is extremely robust and it has a good distinguishability for similar object matching.

Chapter 3. Combination of ABC with correlation filter (ABC-CF)

Our proposed method ABC shows the robustness in the field of image matching at before's chapter. It can effective to deal with the case of rotation, deformation, occlusion, and scale variation [91, 92]. However, ABC is not enough in positional precision due to the loss of pixel location information. Higher sensitivity and positioning precision are needed in some applications, such as matching, pedestrian tracking, image retrieval, and image registration. As well, robustness is also necessary for adverse conditions. As a trial in this paper, these requirements can be realized by combining ABC and another registration scheme of higher positional precision. Therefore, ABC-CF is proposed against the problem of precision.

CF uses the correlation in signal processing to train the filter by extracting target features to filter the input image. This is an effective strategy to enhance matching precision in the process of training and filtering, where color images are transformed into the Fourier domain [93, 94] to reduce the computation cost. Fourier transform is a way of analyzing signals. It can analyze the components of the signal, and it can also use these components to synthesize the signal. Many waveforms can be used as signal components, such as sine waves, square waves, sawtooth waves. The Fourier transform uses a sine wave as the signal component. In the study of image processing, we can use the image as a signal. In Fig. 3.1, p as a reference image is shown in image and

(a) Reference image p in image domain



(b) Reference image p in Fourier domain

Figure 3.1: Reference image p in different domains.

Fourier domains. Then, the sharp peak is produced in the correlation output to obtain the accuracy of the localization of matched images.

In the following section, we introduce how to combine ABC and CF in detail. Among them, it includes the step of the training stage, filter generation modeling, integration of ABC and CF, and the experiments of ABC-CF.

3.1 Training stage of CF

To better overcome the problems of rotation and deformation in the matching process, the input image should first be trained, that is, the reference image. By training the reference image, CF can be less sensitive against rotation or deformation to match the correct target image. It is suitable for combining with robust but rough approaches, such as ABC. Let a reference image be p in image domain. Several p_i are trained as a training set via the affine transformation of the reference image as shown in Fig. 3.2. i is the

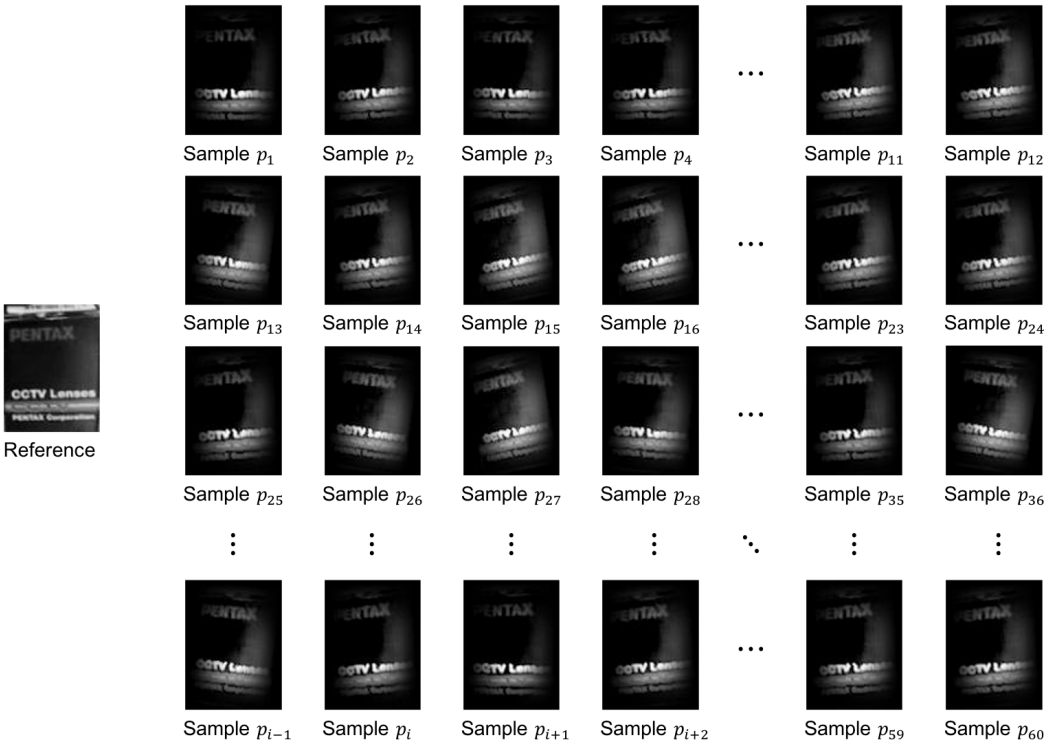
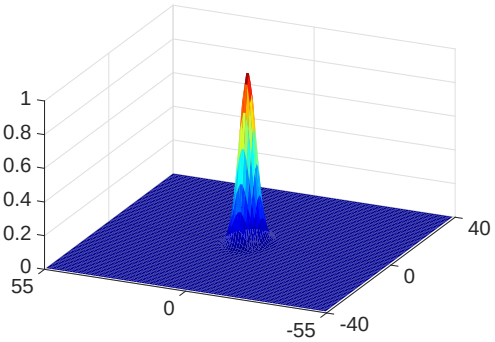


Figure 3.2: Affine transformation for getting reference samples p_i.

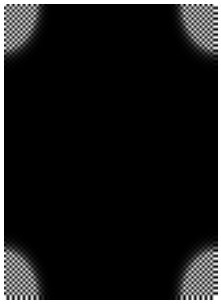
number of samples from 1 to 60 in our experiment. In many applications, the process of training images is involved. Good training can learn the feature information extracted from the image more effectively, but the demerit is that the training is time-consuming. Therefore, it is equally important to establish a reasonable training sample set through experiments.

3.2 Filter generation modeling

Subsequently, an output q in different domains is calculated to use the model of Gaussian-like profile, where $\sigma = 8$, as shown in Fig. 3.3.



(a) Output q with two-dimensional peak-centered Gaussian-like distribution



(b) Output q in Fourier domain

Figure 3.3: The model of output q.

In this method, an optimal filter is defined by a provided two-dimensional peak-centered Gaussian-like distribution and is obtained through several training processes to get a maximum value in a response map, which indicates the best-matched position. The correlation operation with filter u in the image domain was performed via pixel-wise calculations, which is efficiently performed in the frequency domain [95] as follows:

$$Q = P \odot U^* \tag{3.1}$$

where P, Q, and U are the Fourier transforms of p, q, and u, respectively; symbols "$*$" and "$\odot$" indicate the complex conjugate and Hadamard product [96]. To obtain a better filter U, q_i as the output was generated to make a two-dimensional peak. This training is a key process for achieving high and stable sensitivity in finding any precise position in spite of the ill conditions. The minimization of the output sum of squared error (MOSSE) [97] is utilized.

$$\varepsilon = min \sum_i |P_i \odot U^* - Q_i|^2 \tag{3.2}$$

A closed-form expression of U^* is obtained as follows:

$$U^* = \frac{\sum_i Q_i \odot P_i^*}{\sum_i P_i \odot P_i^*} \tag{3.3}$$

3.3 Integration of ABC and CF

Let an image t be the input to CF from the searched position by pre-processing, ABC, as shown in Fig. 3.4. The optimal filter U^* is applied to T, the transformed version of t,

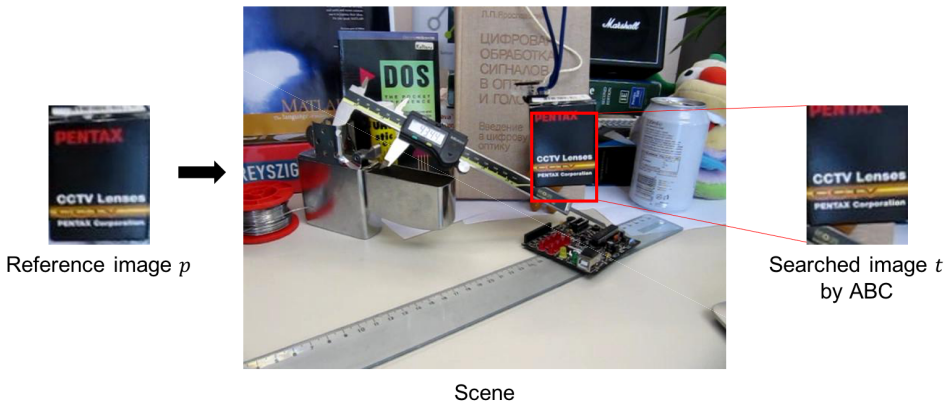


Figure 3.4: Pre-processing by using the ABC approach

for making its response map R in the Fourier domain, in which the largest peak indicates any target position.

$$R = T \odot U^* \tag{3.4}$$

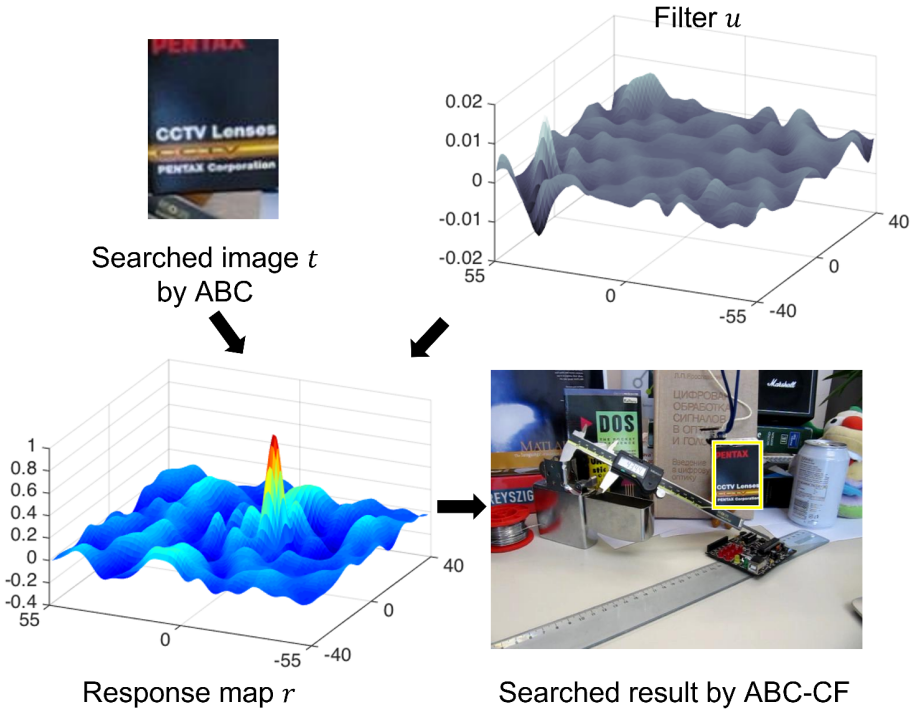


Figure 3.5: Overview of ABC-CF matching.

Fig. 3.5 is an overview of the combination of ABC and CF. As shown in the upper left corner of the figure, ABC as the first step for coarse matching gives a searched image t for initial candidate, and then CF is performed as the next step, in which the filter u in the upper right corner allows for more accurate registration. The figure shows the profile of the response map r in the upper right and the yellow bounding box shows the best-matched position by use of the image data from *Box* dataset [86].

Algorithm 2: Proposed ABC-CF approach

Input: Reference image S_{R} and compared image S_{S}.

Output: Target location L_{T} in the scene.

Initialization: $\sigma = 8$

for *each* $S_{\mathrm{S}}(i,j)$ **do**

- Crop the compared image S_{S} from position (i,j) in the scene.
- Generate two-dimensional color histograms H and G by a* and b* channels.
- Divide color histograms into apparent color histograms ${}^{\mathrm{AP}}H$, ${}^{\mathrm{AP}}G$ and absent color histograms ${}^{\mathrm{AB}}H$, ${}^{\mathrm{AB}}G$.
- Invert absent color histograms ${}^{\mathrm{AB}}H$ and ${}^{\mathrm{AB}}G$ to ${}^{\mathrm{AB}}\bar{H}$, ${}^{\mathrm{AB}}\bar{G}$.
- Normalize apparent and absent color histograms $\left\{H^{\mathrm{AP}}, G^{\mathrm{AP}}, H^{\mathrm{AB}}, G^{\mathrm{AB}}\right\}$.
- Calculate similarity $R_{(i,j)}$ by H^{AP} and G^{AP}, H^{AB} and G^{AB}.

All locations are scanned, then find matched image S_{T} with $max(R_{(i,j)})$.

Training reference image S_{R} by affine transformation for getting samples p_n, where n is from 1 to 60.

Generate outputs q_n with two-dimensional peak-centered Gaussian-like distribution.

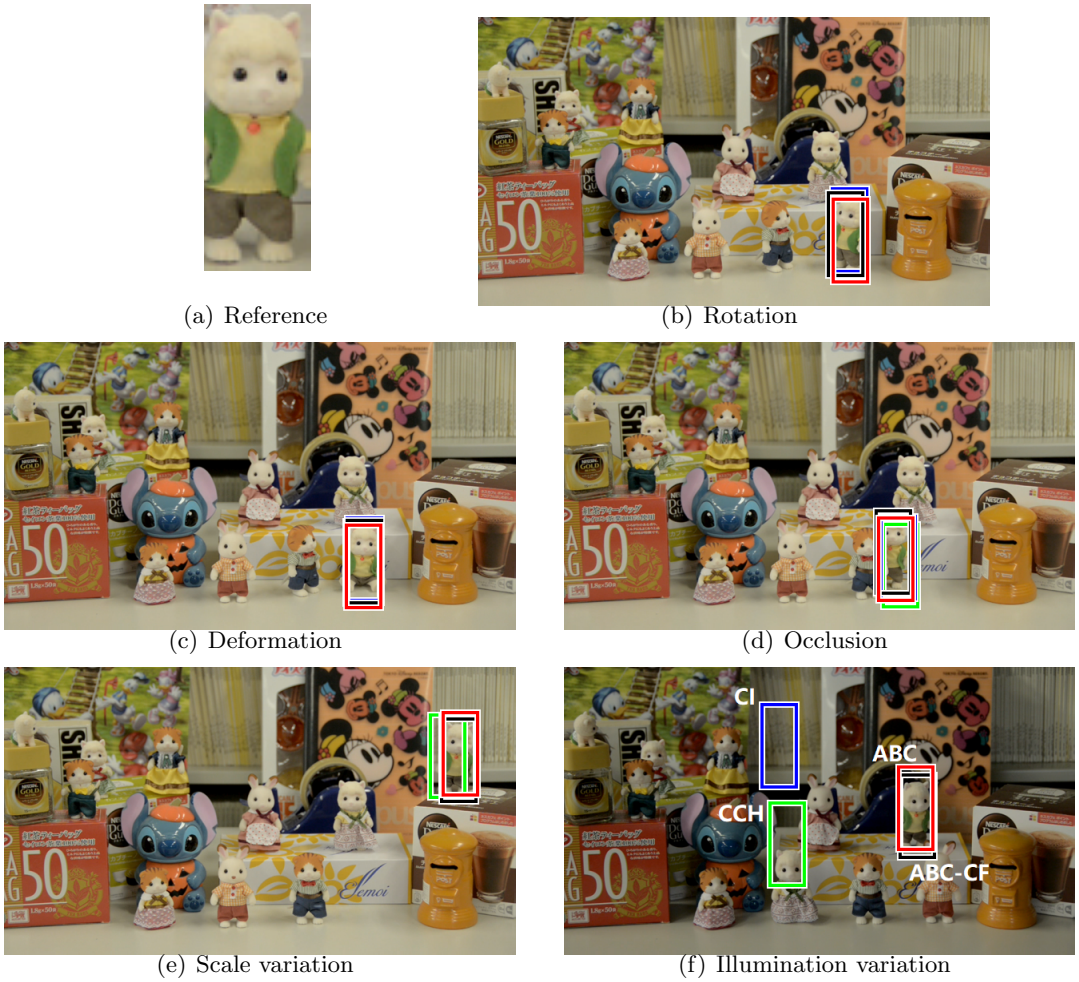
Obtain the correlation filter U^* in the Fourier domain by inputs p_n and outputs q_n.

Calculate the response map by filter U^* and matched image S_{T} from ABC.

Find the target location L_{T} by the maximum value in the response map.

3.4 Performance verification

Some experiments are performed to demonstrate the performance of our proposed method by comparing it with other approaches. In Section 3.4.1, four color histogram-based approaches, i.e., CI, CCH, ABC, and ABC-CF, are compared using real-world images. In Section 3.4.2, not only color histogram-based approaches as comparison, but also template matching approaches are compared for tracking using the open data.

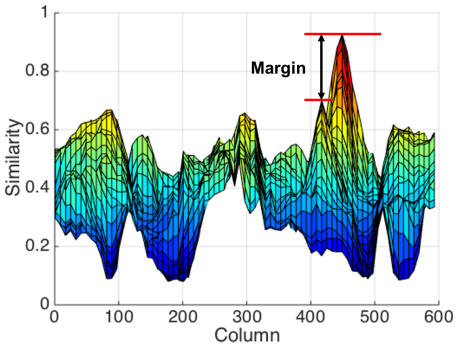


(a) Reference (b) Rotation (c) Deformation (d) Occlusion (e) Scale variation (f) Illumination variation

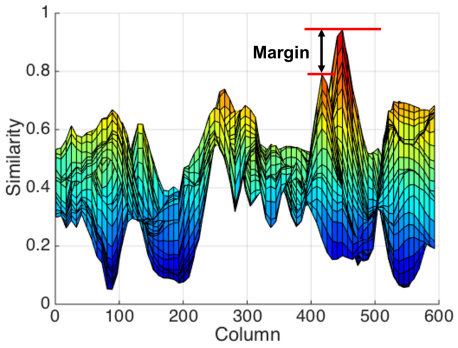
Figure 3.6: Matched results in (a) reference image. (b–f) show matching results by CI, CCH, ABC, and ABC-CF. Bounding black, red, blue, and green boxes show matching results by ABC-CF, ABC, CI, and CCH, respectively.

3.4.1 Search in a cluttered scene

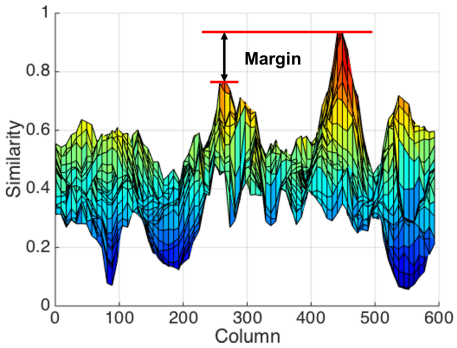
Meanwhile, ABC-CF was selected as an improved version of ABC to compare the matching results. In the experiments, different challenges could be tried, such as rotation, deformation, occlusion, scale variation, and illumination variation, to prove the merits of ABC and ABC-CF. The results obtained in a scene measuring 360×640 are shown in Fig. 3.6. The key feature of ABC is to complete image matching via a color histogram. Therefore, it is compared with some existing color histogram-based methods to evaluate the performance of our approach. Fig. 3.6(a) shows a reference image measuring 100×40. Fig. 3.6(b) through Fig. 3.6(f) show the different challenges to search for the reference position. For the case of rotation, deformation, and occlusion, the methods of



(a) Similarity profile in ABC



(b) Similarity profile in CI



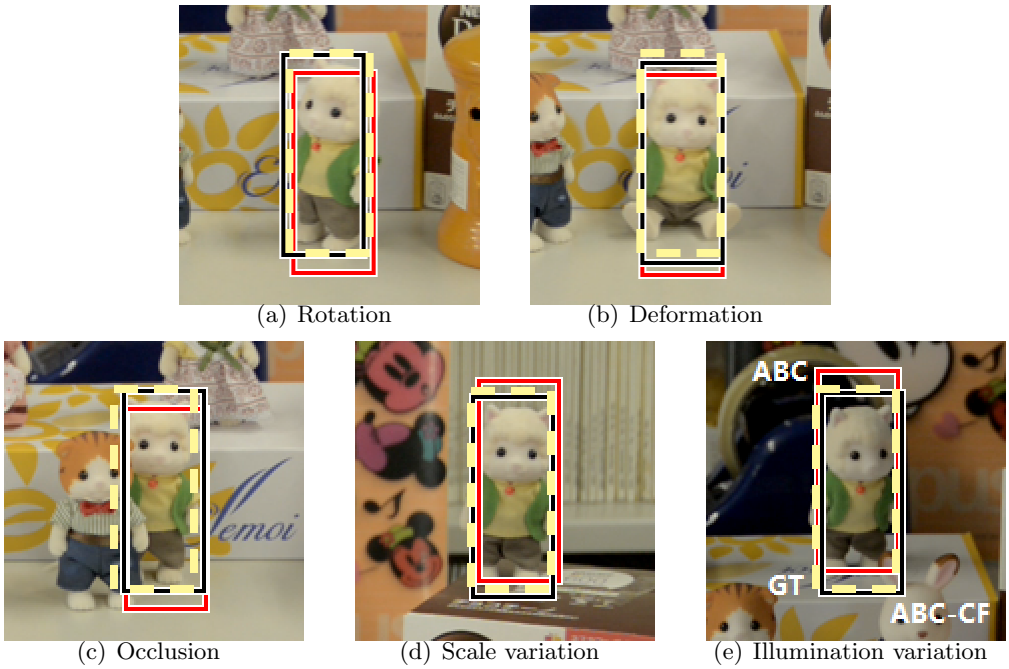
(c) Similarity profile in CCH

Figure 3.7: (a)–(c) show profiles of their similarity in the case of rotation.

ABC, CI, and CCH yielded good performances in the experiments; this demonstrates the advantages of the color histogram-based approaches. In the case of scale variation, the CCH indicated a slight shift, whereas ABC and CI maintained the correct matching position. In the experiment pertaining to illumination variation, only ABC matched with the target, although the matching target position shifted upward slightly.

Table 3.1: Location error comparison with different challenges for color histogram-based approaches.

	CI	CCH	ABC	ABC-CF
Rotation	3.79	**2.23**	9.19	3.16
Deformation	**2.82**	6.09	9.16	5.09
Occlusion	12.04	21.63	10.66	**4.46**
Scale variation	3.41	12.16	3.70	**3.16**
Illumination variation	190.96	164.76	9.05	**3.60**



(a) Rotation (b) Deformation (c) Occlusion (d) Scale variation (e) Illumination variation

Figure 3.8: High precision matching by ABC-CF. Black bounding boxes show ABC-CF matching, while red boxes are ABC results. Yellow boxes are their ground truth (GT).

Fig. 3.7(a)–(c) show the similarity profiles. The best-matched position is compared with the second-best matched position; ABC demonstrated better discrimination ability compared with the other two methods. It was evident that the margin distance of ABC was larger than those of CI and CCH. Fig. 3.8 shows the results of ABC and ABC-CF. ABC-CF, which is the improved version of the original ABC, yielded more accurate searching results and solved the shift problem. GT represents the ground truth for evaluating the performance of the comparison methods. Table 3.1 shows a comparison of the location error based on different challenges for color histogram-based methods, where the location error was calculated based on the Euclidean distance that used GT to compare with the searched position. In cases involving rotation and deformation, the CCH and CI can search for the best position in the experiments. The ABC matching position exhibited a slight downward shift, and this problem was mitigated using the ABC-CF method. In another three cases, the ABC-CF method proved to be the best method as it yielded the lowest location error.

3.4.2 The matching performance of ABC-CF

To evaluate the performance of our new approach, ABC-CF, with color-histogram- and template-based matching methods, the four histogram-based algorithms are selected, i.e., ABC, CI, CCH, and TFCM and two template-matching algorithms, i.e., SSD and NCC, for comparison using open data. Fig. 3.9 shows a reference image of the 85×77

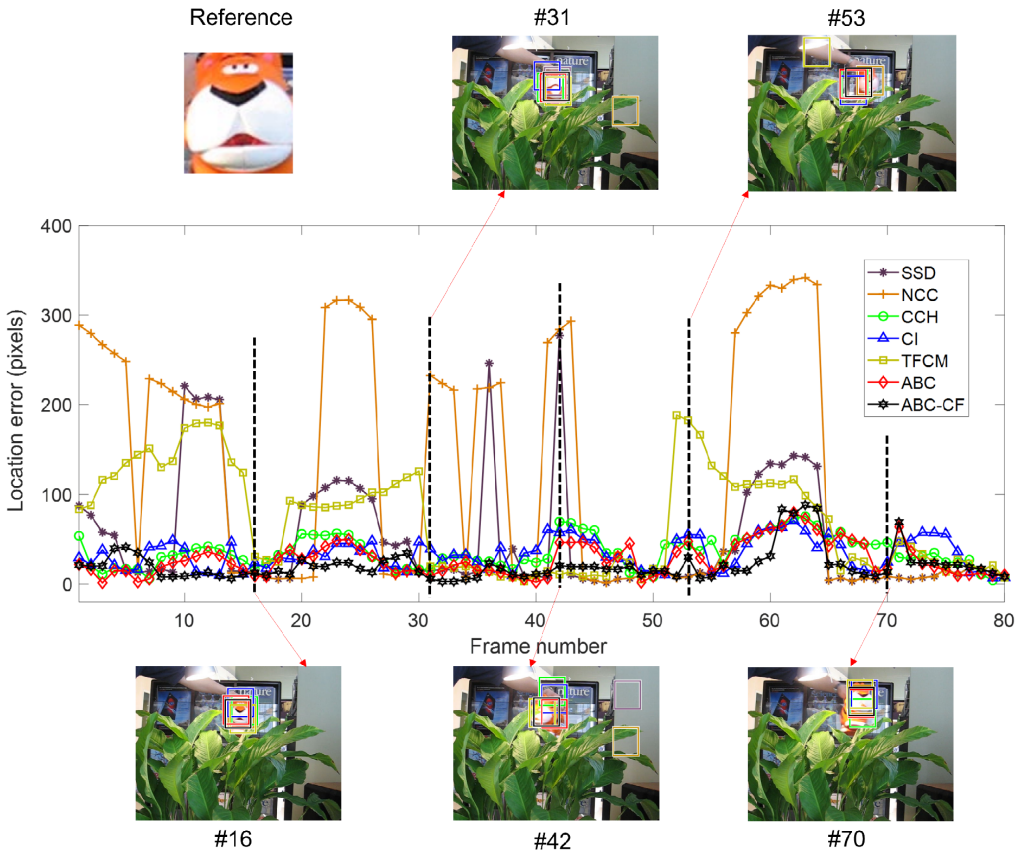


Figure 3.9: Comparison of template- and color-histogram-based methods.

from *Tiger*1 data in [86], where many frames included various instances with severe ill-conditions such as out-of-plane rotation, occlusion, and scaling in many frames. Pixel-by-pixel scanning is done over the scene using the reference image for all the algorithms. In Fig. 3.9, the horizontal axis represents frame label, and the vertical axis represents the location error in the Euclidean distance between their best-matched positions and the ground truths. The five frames are extracted as examples to show some details in finding or matching the reference in the scene. For instance, Frame #31 shows the matching result under the conditions of deformation and illumination variation, where ABC, CI, CCH, TFCM, and SSD obtained better positions despite being slightly shifted. The ABC-CF method yielded the best-matched precision. Similar results were observed in other frames.

The precision plot is shown in Figure 3.10, in which the horizontal axis shows the upper limit of the location error. For example, the precision value at limit 15 signifies the total rate of frames in which the detected positions do not exceed 15 relative to all the frames. The vertical axis shows the precision in the range of 0 to 1. Because both the template-matching-based algorithms failed to increase their precisions as the limits increased, they might have a clear bound of registration up to a distance of approximately

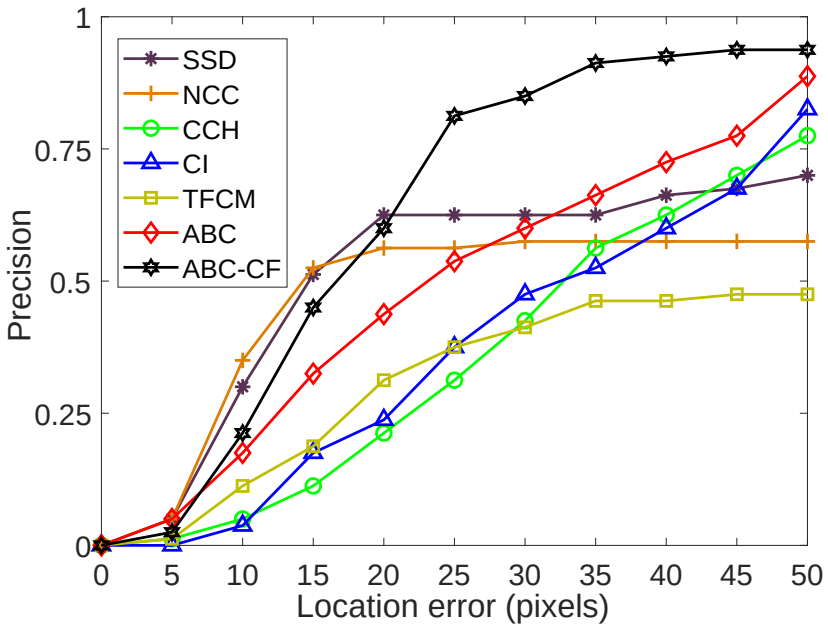


Figure 3.10: Precision plot in *Tiger*1 data.

20 pixels. All the histogram-based approaches did not exhibit such characteristics; however, the precisions are lower than the template-based methods, particularly in the low limits. Table 3.2 shows the number of matching precision in 80 frames. The ABC-CF method demonstrated the best overall performance among all the methods, as indicated by the following findings: higher values around the low limits showed more precise sensitivity in terms of registration performance, whereas higher values in the high limits indicated more robustness identifying the targets.

Table 3.2: Matching precision in 80 frames

	SSD	NCC	CCH	CI	TFCM	ABC	ABC-CF
$Precision \leq 10$	24	**28**	4	3	9	14	17
$Precision \leq 20$	**50**	45	17	19	25	35	48
$Precision \leq 30$	50	46	34	38	33	48	**68**
$Precision \leq 40$	53	46	50	48	37	58	**74**
$Precision \leq 50$	56	46	62	66	38	71	**75**

3.5 Summary

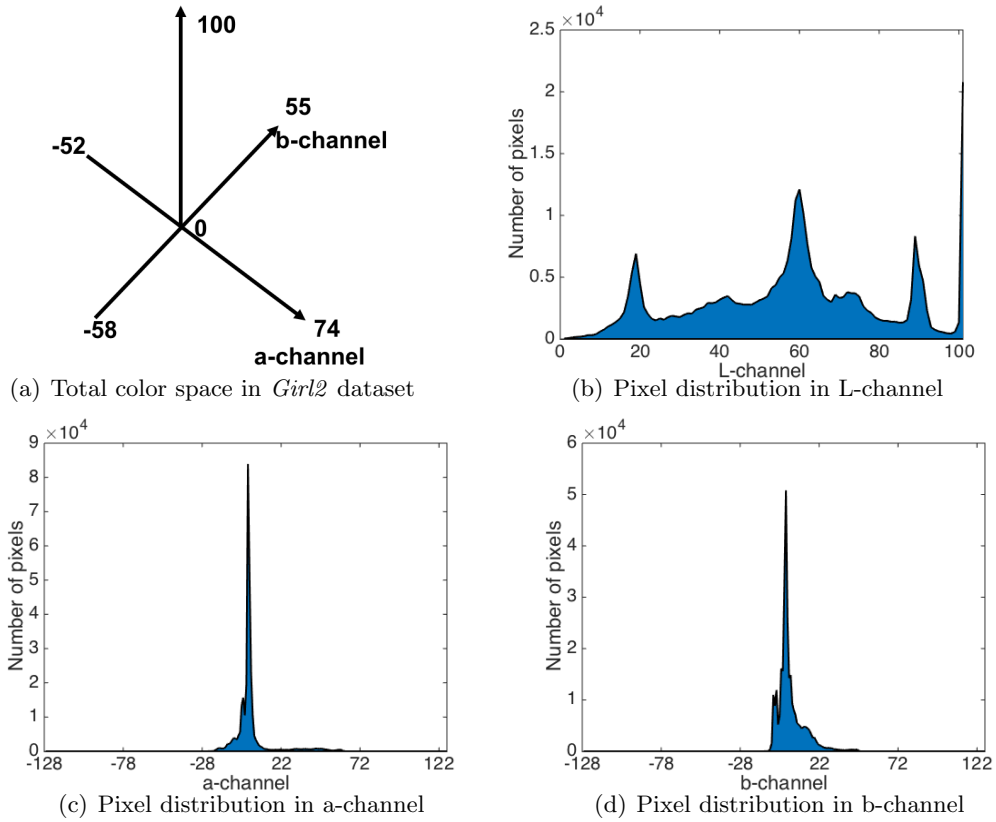
In the chapter, we extend our original idea ABC approach to combine with CF for image matching called ABC-CF. ABC-CF can robust and precise against the demerit of color histogram-based approaches. Firstly, we train the reference image as input and then generate two-dimensional peak-centered Gaussian-like output. After the MOSSE function, a filter u is obtained to do the next step. Finally, we get the matching result of ABC-CF by observing the peak value in the response map. The experiments show the performance of our ABC-CF.

Chapter 4. Improvement of ABC performance by using Multiple-Layered matching (ABC-ML)

Improvement ABC by using the Multiple-Layered structure is proposed called ABC-ML. In the Multiple-Layered model, we implement an effective three layers structure based on the isotonic principle to control the location of each layer. We match each layer by ABC, which is a color histogram-based matching method. Color histogram is decomposed by the ABC approach into apparent and absent color histograms. Low-frequency and non-existent colors are attended to enhance the discrimination ability in the matching process. Next, the Multiple-Layered model is a power supporting against the problem of the offset or drift in color histogram-based methods. Different from ABC-CF in the previous chapter, the ABC approach is integrated with the ML structure. The matching result does not depend on the initial matching result of ABC, but the position information of the color is added in the ABC matching process. Therefore, ABC-ML is an approach that combines local and global color features.

4.1 Total color space

In the field of computer vision [88, 89, 90], numerous color spaces are explored to utilize achieve different algorithms. Generally, we can generate a one- or multi-dimensional color histogram by the range of each component of the selected color space. However, the range of some color spaces far exceeds the range perceived by human vision and selected datasets, such as CIE L *a *b* and other color spaces. Therefore, we define the total color space in a given color histogram as shown in Fig. 4.1. In Fig. 4.1, the *Girl2* dataset [86] as an example is utilized to train the total color space (TCS) in CIE L *a *b* color space. Obviously, the range of a* and b* channels are (-52, 74) and (-58, 55) that is not (-128, 127) and (-128, 127) in *Girl2* dataset. Just the L-channel keeps the same range from 0 to 100. Color histogram-based methods are a statistical way to judge the color feature in images. To calculate the accuracy range of each channel can get a more effective analysis of color feature and reduce the computation time. Even in real-time projects, it is easy to update the TCS frame by frame.



(a) Total color space in *Girl2* dataset

(b) Pixel distribution in L-channel

(c) Pixel distribution in a-channel

(d) Pixel distribution in b-channel

Figure 4.1: Analyze the distribution of pixels on each channel to generate a TCS.

4.2 The introduction of ML structure

Due to the histogram-based matching algorithms are focused on the statistical distribution of given data, hence, it has good properties against rotation, deformation, and scale variations. As a result, it also reduces the matching precision. On the contrary, the template matching algorithms are based on pixels or pixel blocks that have well-controlled accuracy problems. However, there are still weaknesses to search target with the deformation and scale variations produced during the movement process. Fig. 4.2 shows the demerit of color histogram-based methods and the merit of combining with Multiple-Layered matching. For the conventional color histogram-based methods, the matching results in position (a) or (b) have the same similarity. It is because the color distribution of images in positions (a) and (b) are the same, which is why the conventional color histogram-based matching methods have matching accuracy problems. It is difficult to combat the matching offset problem in image processing. In the same situation, combining the merits of color histogram-based method by layering the image can solve the problem caused by the use of the histogram algorithm alone. Multiple-layered matching is based on the isotonic principle to keep the center location is not changed as shown in Fig. 4.3.

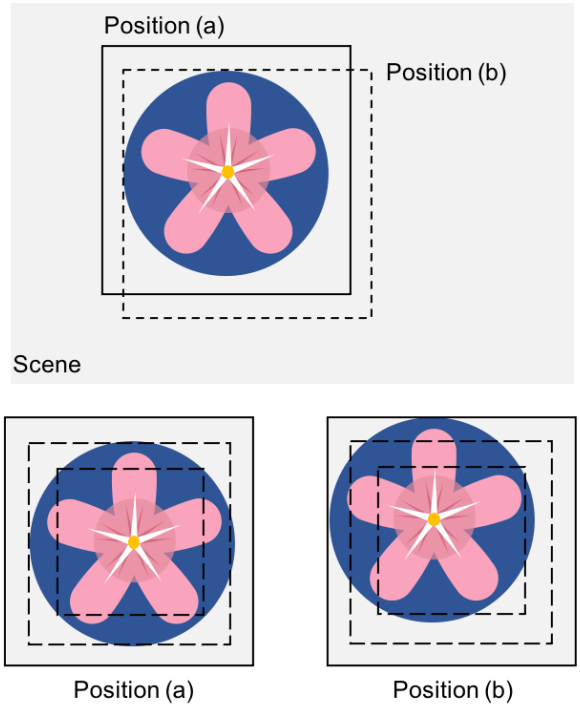


Figure 4.2: The distribution of target color features in different positions under ML matching.

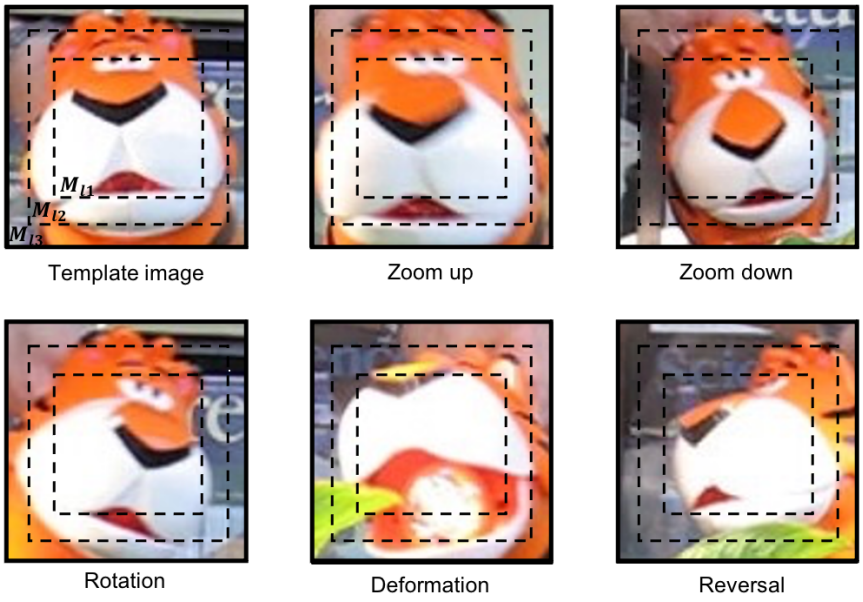


Figure 4.3: The principle of ML matching.

4.3 Combination of ABC and ML

We can observe that when the template image occurs scale variation, rotation, and deformation, the image features at the center zone are essentially kept, especially color features. Therefore, we aim to accurately position the target by dividing the image into multiple layers, and then each layer plays a role in restraining each other's positional relationship to get the optimized matching effect. In Fig. 4.3, $M_{\mathrm{l}1}$ is that the inner layer consists of $T_{\mathrm{m}1}$ pixels in the rectangle; $M_{\mathrm{l}2}$ and $M_{\mathrm{l}3}$ are the second and third layers where they have the relationship as $T_{\mathrm{m}2} = 2 \times T_{\mathrm{m}1}$ and $T_{\mathrm{m}3} = 3 \times T_{\mathrm{m}1}$.

Next, we assume that the position P_{m3} in Eq. 4.3 and size of the template image are [0, 0] and [h, w], where the initial position of the image [0, 0] is at the upper left of the image. Following the given two conditions, first, the center point of each layer is consistent; Second, the relationships between each layer of pixels are $T_{\mathrm{m}2} = 2 \times T_{\mathrm{m}1}$ and $T_{\mathrm{m}3} = 3 \times T_{\mathrm{m}1}$. Thereafter, the position L_{m2} and size of template image in $M_{\mathrm{l}2}$ layer is in Eq. 4.2, and the L_{m1} in layer $M_{\mathrm{l}1}$ is in Eq. 4.1.

$$L_{\mathrm{l}1} = \left[\left(1 - \frac{\sqrt{6}}{3}\right) \times \frac{h}{2}, \left(1 - \frac{\sqrt{6}}{3}\right) \times \frac{w}{2}, \frac{\sqrt{6}}{3}h, \frac{\sqrt{6}}{3}w \right] \tag{4.1}$$

$$L_{\mathrm{l}2} = \left[\left(1 - \frac{\sqrt{3}}{3}\right) \times \frac{h}{2}, \left(1 - \frac{\sqrt{3}}{3}\right) \times \frac{w}{2}, \frac{\sqrt{3}}{3}h, \frac{\sqrt{3}}{3}w \right] \tag{4.2}$$

$$L_{\mathrm{l}3} = [0, 0, h, w] \tag{4.3}$$

The similarity $S_{i=1,2,3}$ of each layer is obtained by using the absent color indexing method, Multiple-Layered similarity S_{ml} is expressed as follows

$$S_{\mathrm{ml}} = \frac{1}{3} \times \sum_{1}^{3} S_i \tag{4.4}$$

4.4 Effectiveness of ABC-ML for image matching

The proposed ABC-ML can aginst the shift problem in color histogram-based methods without combining with other features. First, we implement experiments with color indexing (CI), cumulative color histogram method (CCH), and absent color indexing (ABC) to validate the performance of the proposed ABC-ML method in Section 4.4.2. Next, we evaluate ABC-ML with template-based matching methods, such as SSD, NCC, BBS, to demonstrate the effect of the proposed robustness and accuracy of ABC-ML in Section 4.4.2.

Algorithm 3: Proposed ABC-ML approach

Input: Reference image S_{R} at position L_3 and compared image S_{S}.
Extract each layer M_{κ}^{r} of image S_{R} and M_{κ}^{s} of image S_{s}.
Train the given dataset to determine the range of TCS.

Output: Target location L_{T} in searched image.

Initialization: $\kappa = 3$.
for *each* $S_{\mathrm{s}}(i,j)$ **do**
  Extract an image S_{s} at position (i,j) in the scene.
  for $\kappa > 0$ **do**
    Extract the layer M_{κ}^{r} of image S_{R} and M_{κ}^{s} of image S_{s} in position L_{κ}.
    Generate two-dimensional color histograms H_{κ} and G_{κ} by images M_{κ}^{r} and M_{κ}^{s} in TCS.
    Divide color histograms into apparent color histograms H_{κ}^{AP}, G_{κ}^{AP} and absent color histograms ${}^{\mathrm{AB}}H_{\kappa}$, ${}^{\mathrm{AB}}G_{\kappa}$.
    Invert absent color histograms ${}^{\mathrm{AB}}H_{\kappa}$ and ${}^{\mathrm{AB}}G_{\kappa}$ to ${}^{\mathrm{AB}}\bar{H}_{\kappa}$ and ${}^{\mathrm{AB}}\bar{G}_{\kappa}$.
    Normalize apparent and absent color histograms to $\{H_{\kappa}^{\mathrm{AP}}, G_{\kappa}^{\mathrm{AP}}, H_{\kappa}^{\mathrm{AB}}, G_{\kappa}^{\mathrm{AB}}\}$.
    Calculate similarity $S_{\kappa(i,j)}$ by H_{κ}^{AP} and G_{κ}^{AP}, H_{κ}^{AB} and G_{κ}^{AB}.
    $\kappa = \kappa - 1$
  Similarity $S_{(i,j)} = \frac{1}{3}S_{1(i,j)} + \frac{1}{3}S_{2(i,j)} + \frac{1}{3}S_{3(i,j)}$.
All locations are scanned, then find position L_{T} with $max(S_{(i,j)})$;

4.4.1 Analysis of performance

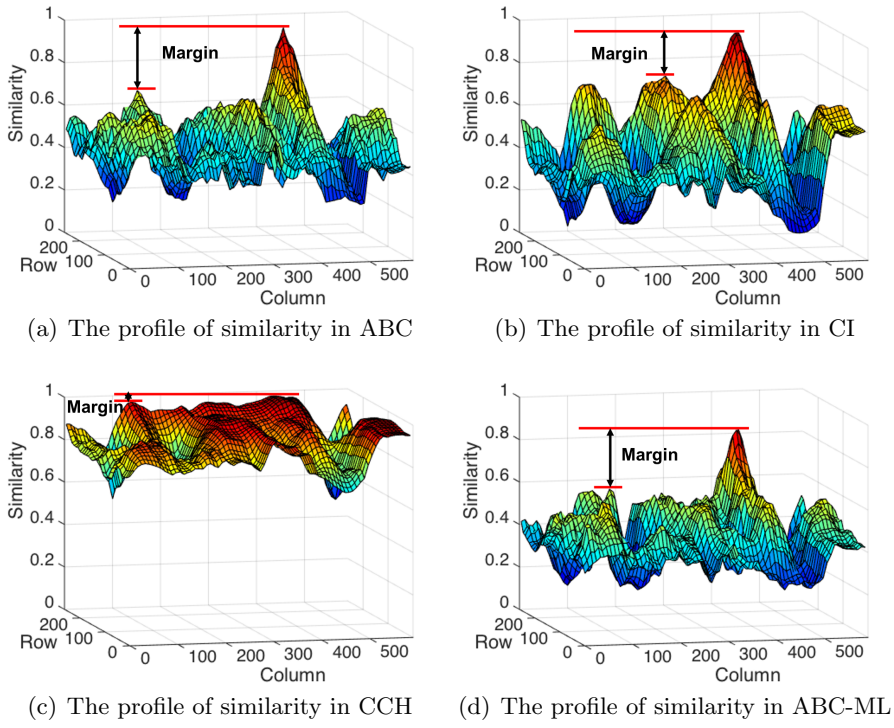
To fair comparing, we set using experimental parameters which are identical to 2 for all images or sequences in ABC-ML. We evaluate the proposed ABC-ML method on real-world images. These images are analyzed with five challenges for detailed analysis, including rotation, deformation, scaling, occlusion, and illumination variation. We set the reference image in Fig. 4.4(a) where the size is 98×50. The scenes are the same size 360×640 in the experiments. We compare the best-matched position from each color histogram-based method with ground truth, thereafter, calculate the Euclidean distance to evaluate the experimental results. By observing Fig. 4.4, we can find color histogram-based methods have a property on robustness. But, it is unstable for the precision during the matching process. Therefore, ABC-ML is proposed to improve ABC for handling the matching precision problem.

All of the four methods could correctly find the position of the reference image although under the case of rotation, but in observation of their profiles, we could see that the difference between the highest peak and the second peak of in ABC- and ABC-ML-based search were larger than the other two as shown in Fig. 4.5. Margin is a symbol that has a better discrimination ability among these classes. Fig. 4.6 shows the matching results between ABC and ABC-ML. Ground truth presents by white. We observe that the

(a) Reference (b) Rotation (c) Deformation (d) Occlusion (e) Scale variation (f) Illumination variation

Figure 4.4: Matching results in color histogram-based methods. (a) is a reference image. (b)-(f) show matching results by CI, CCH, ABC, and ABC-ML under different challenges. Bounding blue, green, red, and black boxes show matching results by CI, CCH, ABC, and ABC-ML, respectively.

results in ABC can find roughly location but it is not accurate. ABC-ML is more near ground truth. In ABC-ML, the center point in each layer is not changed and improves the matching accuracy through the positional relationship and mutual control of each layer. Table. 4.1 shows a comparison of the location error based on different challenges for color histogram-based methods, where the location error was calculated based on the Euclidean distance that used the ground truth to compare with the searched position. In cases involving rotation and occlusion, the CI and ABC can search for the best position in the experiments. In another three cases, the ABC-ML method proved to be the best method as it yielded the lowest location error.



(a) The profile of similarity in ABC

(b) The profile of similarity in CI

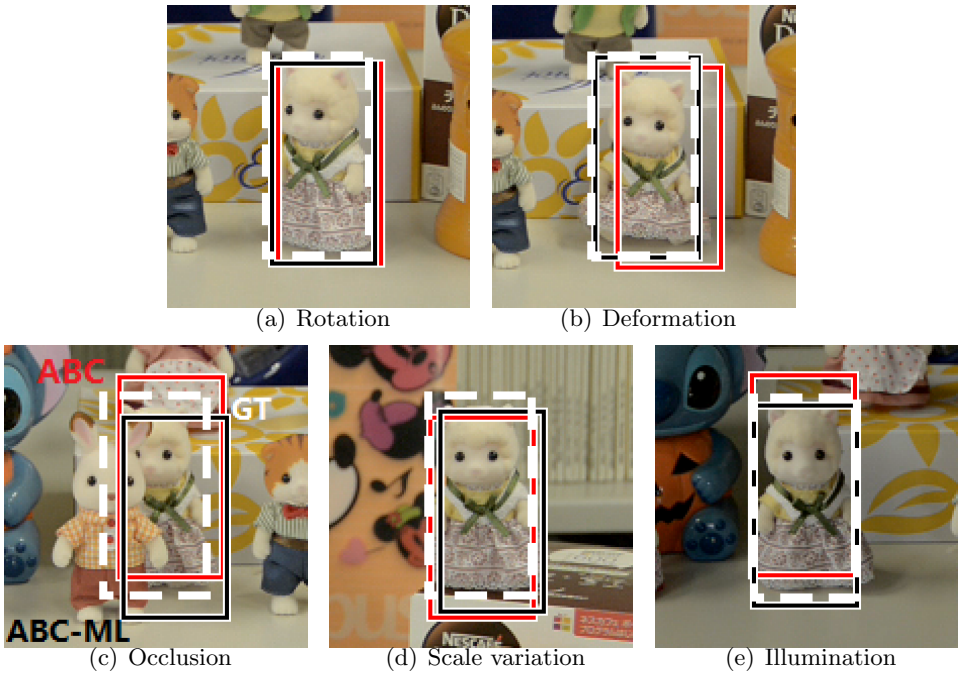
(c) The profile of similarity in CCH

(d) The profile of similarity in ABC-ML

Figure 4.5: Profile plots which are used CI, CCH, ABC, and ABC-ML methods in Fig. 4.4(b)

4.4.2 Robustness of ABC-ML matching

We compare our approach with three color histogram- and three template-based matching methods, including ABC, CI, CCH, SSD, NCC, and BBS. We select open data for comparison and present the results using location error that is Euclidean distance between the searched center- and ground truth center-point. *Skiing* data in [86] is utilized to evaluate the performance of six methods, where the size of reference image is 33×36. In Fig. 4.7, the horizontal axis represents the frame label, and the vertical axis represents the location error in the Euclidean distance between their best-matched positions and the ground truths. Some representative frames as examples are shown in detail. Frame #35 shows the good matching result in six methods, and only SSD can not find the skier. In Frame #42, NCC matches the position around the tree, not the target. The template-based matching approach calculates the pixel value; therefore, the similarities between correct and incorrect positions are almost identical. Frames #61 and #64 show the deformation condition if compared with the reference image. ABC and ABC-ML can catch the absent colors, red and yellow, to match the correct position under the case of deformation. The skier is far away from the camera, and background information becomes more large proportion. Hence, another five methods focus on the main colors, which can not match the correct position. Meanwhile, we observe that the



(a) Rotation (b) Deformation

(c) Occlusion (d) Scale variation (e) Illumination

Figure 4.6: Increasing matching precision with ABC-ML. Black bounding boxes are ABC-ML. Red boxes show ABC results. Bounding white boxes are ground truth.

Table 4.1: Location error in color histogram-based methods.

	CI	CCH	ABC	ABC-ML
Rotation	**3.60**	5.09	7.21	4.24
Deformation	3.60	11.04	14.31	**2.23**
Occlusion	29.06	158.1	**12.04**	14.86
Scale variation	11.40	16.27	11.04	**9.21**
Illumination variation	7	4.12	11.04	**4**

precision of ABC-ML is better than our original proposed ABC approach. The ABC-ML approach demonstrated the best overall performance among all the approaches.

4.5 Summary

We exploit a Multiple-Layered structure against the shift problem of matching in color histogram-based methods. It utilizes the isotonic principle to construct each layer. To challenge the various matching problem, we combine Multiple-Layered with ABC (ABC-ML). In the experiments, we evaluate our approach to Mondrian random patterns, real-world images, and open databases, and it obtains state-of-art results. Moreover, the proposed ABC-ML performs superiorly against deformation and scale variances. Experiments are conducted on real-world data and benchmark sequences. ABC-ML can

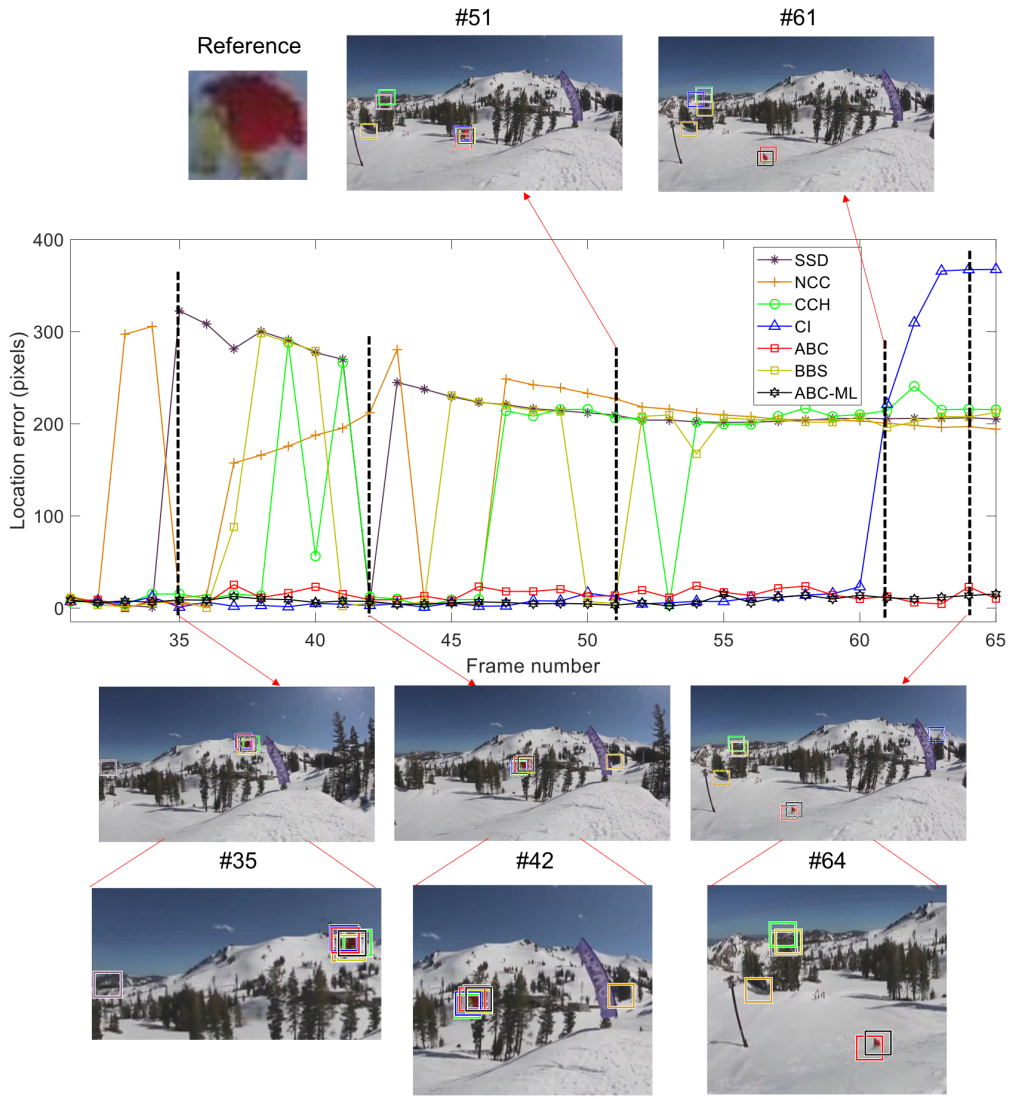


Figure 4.7: Matching results in *Skiing* data.

keep the merit of ABC on a large margin, and also provide the location information of color features by combining with a Multiple-Layered model. Under the different challenges, ABC-ML has stable performance in the area of image matching.

Chapter 5. Experimental evaluation

This chapter introduces the experimental evaluation for ABC, ABC-CF, and ABC-ML approaches, including five main parts: (1) Experimental setup is described for initial experiment parameters. (2) How to select a reasonable threshold h_{T}? (3) Do the comparison experiments by using three of our proposed approaches to explain the merits and demerits for each approach. (4) Computation cost is shown. (5) Discussion will be presented to conclude experimental results for our proposed three approaches. In the final, we summarize this chapter.

5.1 Experimental setup

In the experiments, we add some artificial noises to original images for investigating the performances of our proposed ABC approach. Fig. 5.1 shows an example of a synthetic Mondrian random pattern with no structures except for a colored circle as an elemental shape.

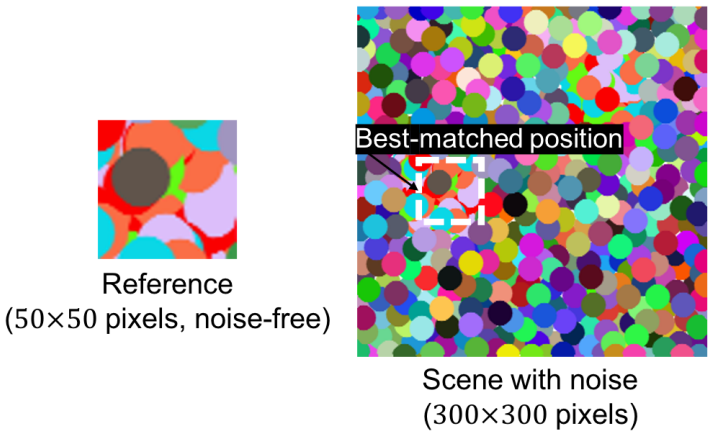


Figure 5.1: Mondrian random pattern with noise.

We tested four similarity measures as follows: First, reference images of the size 50×50 in Fig. 5.1 were randomly selected in noise-free image, and then searching for the best-matched positions were performed in each noisy version. In Fig. 5.2, the abridged general view of the analysis measures is shown, such as *Precision* and *Recall*. To evaluate

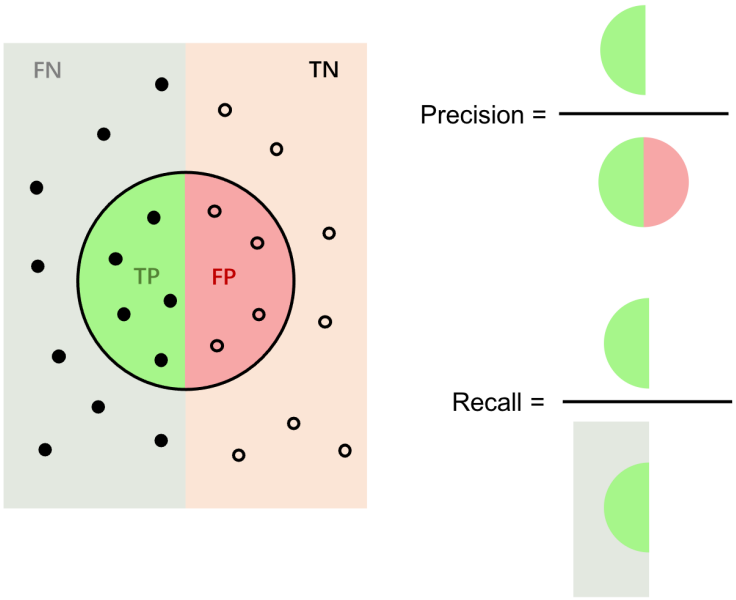


Figure 5.2: Abridged general view of the analysis measurements.

each similarity measure, we use *F-measure* as

$$Precision = \frac{TP}{TP+FP} \tag{5.1}$$

$$Recall = \frac{TP}{TP+FN} \tag{5.2}$$

$$F\text{-}measure = \frac{2 \times Precision \times Recall}{Precision+Recall} \tag{5.3}$$

where TP and FP indicate numbers of true positives and false positives, respectively,

Table 5.1: Performance of ABC with various similarity measures

	F-measure		
SNR	36	33	30
ABC + Intersection	0.973	0.965	0.952
ABC + Chi-square	0.976	0.971	0.961
ABC + JS divergence	0.980	0.974	0.971
ABC + Bhattacharyya	0.974	0.970	0.968

defined by a threshold value of intersection over union (IoU). TP can thus have IoU exceeding 0.75, while FP has a value below threshold 0.25. FN represents the number of false negatives, where $TP+FN$ is the entirety of the target image. *F-measure* is

the harmonic mean of *Precision* and *Recall*. Table 5.1 shows a performance evaluation in matching using ABC-based similarity measures under different noise conditions. We found that in all cases, ABC could be used in combination with these similarity measures.

In this thesie, we set experimental parameters $I = 10$, $J = 10$, $\alpha = 0.2$, $w_{\mathrm{AP}} = 0.6$, and $w_{\mathrm{AB}} = 0.4$. Meanwhile, we delete noise by $0.2 \times h_{\mathrm{T}}$ before decomposing. In ABC-CF, we additional add parameters $\sigma = 8$. We use *Intersection* in the experiments because of its similarity in the range from 0 to 1, which makes it easier to compare with other approaches. The datasets [86] are introduced as follows:

- **Box dataset:** a typical sequence of scale variations. A blue line pulls up a box with black color and moves randomly in the cluttered scene. Sometimes, there will be occlusion and rotation. The static background in this dataset also includes many objects with the same color as the target, such as books and stereos.

- **Liquor dataset:** the sequence with the challenges of rotation, deformation, scale variation, and occlusion. The bottles were picked up using a human finger, moved from one place to another, and exchanged with the location of another bottle. Meanwhile, some bottles are similar in colors. The background is static, and there are no illumination variations.

- **Skiing dataset:** one sequence with strong deformation and rotation is selected for evaluating the performance. A skier wearing a red coat and yellow pants slipped into the air and completed complicated movements such as flips. In the dataset, the scenes are dynamically changing and similar, mainly composed of the colors of trees and snow. For the skier, there have been scale variations from the beginning to the end of the sequence.

- **Girl2 dataset:** a pedestrian tracking sequence. A girl plays in a dynamic and complex scene, which includes various matching challenges. Although there is no obvious light change, the colors in the background are slightly different due to changes in natural light. The appearance of occlusion increases the difficulty of image matching.

- **Tiger1 dataset:** one typical sequence with illumination variations is utilized. The tiger toy with orange color is picked up by a human hand. The tiger toy is sometimes close to the light during movement and sometimes far away from the light. The background is static. In the scene, there is a row of plants in front of the target as occlusion to increase the difficulty of matching.

Above these sequences as experimental datasets are used to evaluate the effectiveness of our proposed ABC, ABC-CF, and ABC-ML. By observing the datasets, we also can find that the above datasets contain five different image matching challenges: rotation, deformation, occlusion, scale variation, and illumination variation.

5.2 Parameter discussion

Threshold h_{T} is an essential parameter in the concept of ABC. How to get a meaningful threshold h_{T} is also a difficult problem that this thesis needs to overcome. In previous version of ABC approach [91, 92], we propose a fix threshold h_{T} to robust matching. To get more reasonable threshold, we define the parameter α to determine how low frequencies are absent colors, and then h_{T} is obtained by mean color histogram. The detailed explanation is described in Section 2.4. Therefore, we tested the effectiveness of introducing α to determine h_{T} using the *Liquor* dataset. As preparation, we scanned all possible positions in frames #1301–#1400 to calculate the average $\bar{h}_{\mathrm{T}}$ by 100 frames' $\hat{h}_{\mathrm{T}}$ and its variance $\sigma^2_{\mathrm{h_T}}$, which were obtained as 0.05 and 6.96×10^{-4} by introducing $\alpha = 0.2$. Fig. 5.3 shows the mean of $\hat{h}_{\mathrm{T}}$ under $\alpha = 0.2$ in each frame, and $\bar{h}_{\mathrm{T}}$ is the average value of $\hat{h}_{\mathrm{T}}$ in these 100 frames. From this result, it may be reasonable

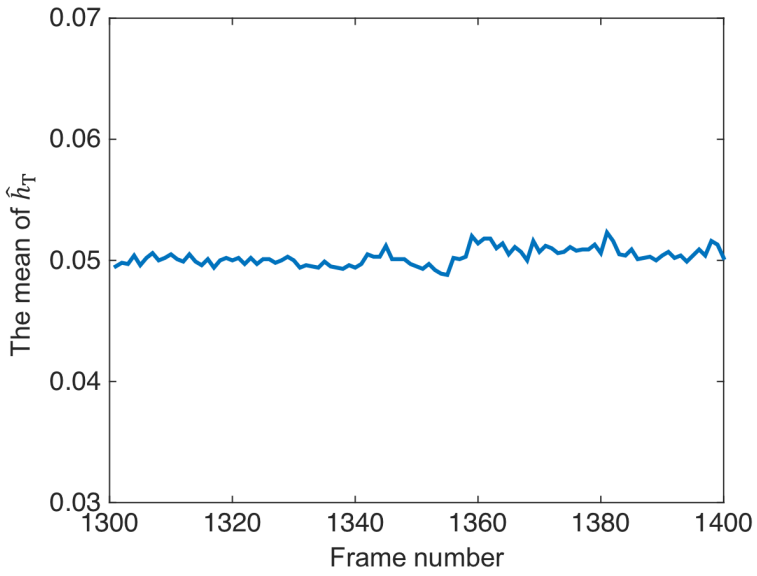


Figure 5.3: The mean of $\hat{h}_{\mathrm{T}}$ under $\alpha = 0.2$ in each frame.

to compare the two cases of $\alpha = 0.2$ and fixed value $h_{\mathrm{T}} = 0.05$, which could be selected as representative constant values.

Fig. 5.4 shows that in these two cases, where the red line is the case of $\alpha = 0.2$ and the black line is the case of fixed value $h_{\mathrm{T}} = 0.05$. The reference image is shown in the top left. Thereafter, the reference image matches the target in the scene from frames #1301 to #1400, respectively. As a result, we could find no difference in performance except for the case of frame #1373, where the constant h_{T} failed to find the correct position. The experiment proves that getting h_{T} by using $\alpha = 0.2$ is stable and accurate. Therefore, it is more robust to update the threshold h_{T} used to separate apparent and absent colors by the color histogram features of the two compared images in image matching. In this

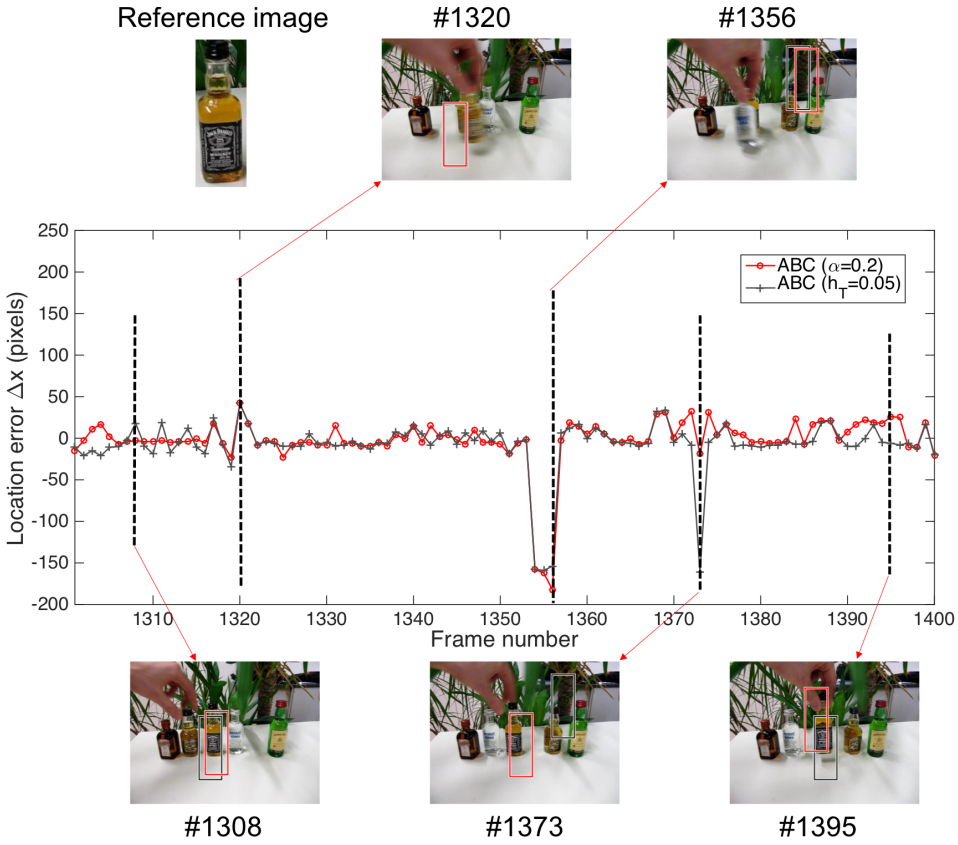


Figure 5.4: Comparison of two cases for threshold selection.

paper, $\alpha = 0.2$ as a parameter is used for the experiments.

5.3 Experimental comparison

To demonstrate the proposed method's ability, ABC, ABC-CF, and ABC-ML, we select the *Girl*2 dataset to do the experiments. This dataset includes different challenges for matching. Next, we will analyze these three approaches through consecutive sequence matching experiments. As can be seen from Fig 5.5, the size of the reference image in the top left is 151×50 pixels. The size of the scene is 480×640. In frames #714, #769, and #773, the girl in the searched image is not large change to compare with the reference image. However, ABC has a litter bit of offset or drift when we compare the matching results with ABC-CF and ABC-ML. The reason is histogram-based information loss the position information. Frame #728 and #743 show the result of ABC-ML; it is not accurate in the matching position because ABC-ML obtains less position information to compare with the template matching algorithm. The merits of these three approaches are (1) absent colors can effective and robust apply in image matching. (2) ABC-CF and ABC-ML improve the precision of algorithms as shown in Fig 5.6. Table 5.2 lists the

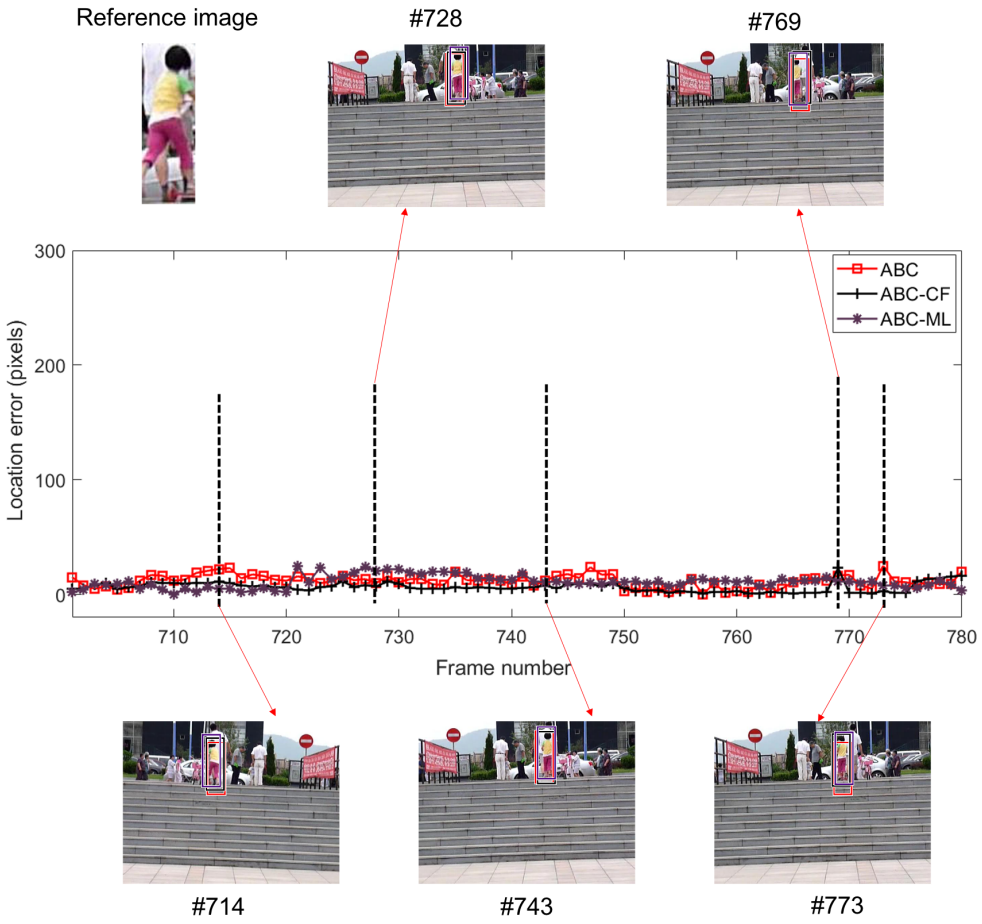


Figure 5.5: The matching performance in ABC, ABC-CF, and ABC-ML.

Table 5.2: Searching precision in 80 frames

	ABC	ABC-CF	ABC-ML
$Precision \leq 5$	11	**26**	14
$Precision \leq 10$	29	**55**	34
$Precision \leq 15$	58	**71**	64
$Precision \leq 20$	75	**79**	75
$Precision \leq 25$	79	**80**	**80**
$Precision \leq 30$	**80**	**80**	**80**

precision evaluation results. ABC-CF has a good performance compared with another

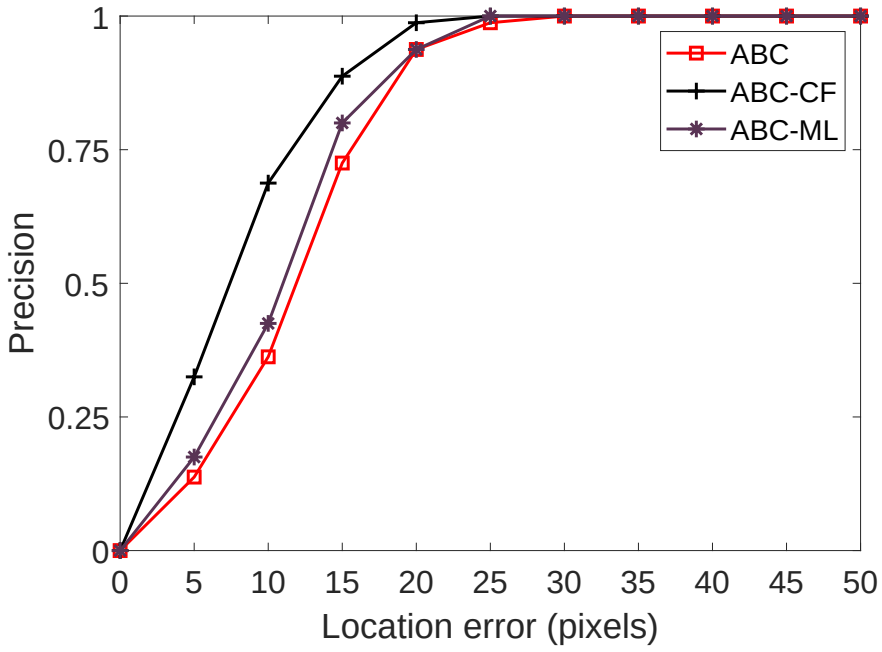


Figure 5.6: The plot of precision.

two approaches. ABC-ML enhances the accuracy of our original ABC as well. The experimental results prove the necessity of our proposed approaches.

5.4 Computation cost

This section compares the processing time in our proposed methods ABC, ABC-CF, and ABC-ML. The programs for the experiments were implemented in C++ by using Visual Studio 2015 and the OpenCV 2.4.13 library, without any parallel processing or GPU acceleration. The hardware was a Windows 10 PC with a 2.81 GHz Intel Core i5-8400 CPU and 8 GB RAM. Since the approach proposed in the paper, ABC, ABC-CF, and ABC-ML, is based on pixel-by-pixel calculation in nature, the computation cost is proportional to the number of pixels in the reference images and the target scene. The matching task depicted in Fig. 5.7 was selected as a typical example to check the computation time requirements. A reference image of 170 $\times$ 90 pixels and a scene of 360 $\times$ 640 were used in this task, and the computation time was then observed using the OpenCV timing function. Table 5.3 shows the computation costs for these three approaches. The most time-saving approach was ABC because the matching processes of ABC-CF and ABC-ML are based on ABC. Although ABC-CF and ABC-ML exhibited a time disadvantage, this is not problematic in practical applications because the computation cost of all histogram-based methods does not differ significantly. In the algorithm based on histogram matching, it is generally necessary to convert the image to another color space

(a) Reference image (b) Scene

Figure 5.7: Example of calculating time consumption.

before generating the histogram. Therefore, it is needed to perform statistical analysis on the information of each pixel in the image. However, with the continuous progress of research and development, the problem of computation cost can be solved by computer hardware. Therefore, time-consuming can be evaluated as an indicator of the algorithm. At the same time, it can also be done by improving the algorithm and the cooperation of the hardware.

Table 5.3: Computation costs for three approaches.

	ABC	ABC-CF	ABC-ML
Computation cost	8.597 s	9.104 s	13.092 s

5.5 Discussion

We summarize some aspects of the proposed technique in this section. ABC has some merits: technical simplicity, invariance with respect to in-plane rotation, distortion, and somewhat scaling as shown in the previous chapter, but it remains some demerits: the necessity of colors and reasonable size of images to keep histograms effective, and loss of positional information. In many of our experiments, we showed its effectiveness in tracking and searching problems. In general, these problems need a rapid method both in offline processing and in online or real-time calculation. When we have enough data of training images, it may be possible to utilize many ways of machine learning technologies, such as CNN or effective classifiers, to solve these problems. Therefore, compared to machine learning, ABC is a simple approach in structure and no need for preprocessing training data. ABC is a histogram-based statistical algorithm to extract color information for image matching. It is unlike deep neural networks; our proposed approach does not require any negative samples. The computation cost of ABC is relatively lower than deep neural networks. However, we hope the cooperation of the proposed technique

with them, for instance, to reduce the size or cost of training and to raise the total performance in processing. For example, in some hierarchical approaches, ABC can be effective to nominate possible candidates for continuing the following detailed classification or some other applications, such as defect detection, pedestrian tracking. In this type of utilization, it must be preferable for absent colors to be somewhat independent of or orthogonal to any other features to keep better performance in total. To upgrade the performance of ABC, we proposed ABC-CF and ABC-ML to solute the problem of loss of positional information. The merit of ABC-CF is that the filtering speed is fast, and the matching position is more accurate. The disadvantage is that CF relies on the rough matching result of the ABC algorithm in the second positioning process. Therefore, if the initial matching result of ABC is incorrect, it will directly cause the relocation of CF to fail. ABC-ML starts from the image itself and structures the image to analyze the color features in the corresponding structure. Compared with ABC-CF, it does not rely on the matching results of our proposed ABC method. It obtains location information through its structure to achieve the purpose of improving matching accuracy. The demerit of ABC-ML is that it is time-consuming. To solve the demerits of our proposed approaches, we will analyze the absent colors to extract effective absent colors for generating an absent color histogram.

5.6 Summary

In the chapter, we first introduced the design of the experimental setup. Then, the significant parameter is discussed to explain why we select $\alpha = 0.2$ to obtain threshold h_{T}. We compare the performance of ABC, ABC-CF, and ABC-ML under different challenges. Experiment results show the effectiveness and robustness of our proposed ABC approach. Meanwhile, improvement ABC-CF and ABC-ML have good properties for solving the offset or drift problem. Next, the computation cost is compared in these three approaches. In the final, we discuss the ABC technique and its possible application.

Chapter 6. Conclusions and future works

6.1 Conclusions

In this paper, we propose a novel method based on color histograms called absent color indexing (ABC) for robust image matching. ABC addresses the constraint of conventional color histogram-based approaches, which focus on the main or prominent colors. Meanwhile, the ABC algorithm increases the importance of minor colors that comprise a relatively small proportion and provides a fair and reliable evaluation for the histogram-based matching algorithm. By reorganizing a color histogram into the two complementary histograms, we were able to observe that a new feature of absent colors is effective in increasing the margins, resulting in high reliability or distinguishability in many different tasks. Subsequently, we decompose the color histogram into two histograms: apparent and absent color histograms. Low-frequency colors or relatively non-existing ones in the color histogram bins are enhanced by inversion for fair treatment in the proposed ABC. A novel method to determine the important parameter for histogram decomposition was provided by specifying the statistical significance level in the mean histogram. Finally, ABC is combined with four similarity measurement methods to calculate similarity.

Furthermore, based on the ABC approach, a correlation filter (CF) is utilized to fuse ABC to address offset or drift cases in process of image matching. The ABC algorithm is used to coarsely locate the target, and the final matching result of the target is performed by the ABC searched position combined with CF. The ABC-CF has improved the matching precision and kept the merit of robustness when we search the target in the scene. Therefore, the highest peak in the response map shows the best-matched position.

As a new concept, the multiple-layered (ML) structure is presented. It was developed to improve matching accuracy compared to the use of only ABC. ABC-ML adds location information to the color information through the relationship between layers, thereby compensating for the insufficiency of the color histogram algorithms.

To verify the performance of our proposed ABC, Mondrian random patterns were effectively used for the fundamental evaluation of the proposed method in comparison experiments with some representative competitors. ABC has good distinguishability and robustness for similar objects. Both ABC-CF and ABC-ML effectively improved the precision of matching locations under different challenges. Experimental results on

the use of real-world images and open datasets showed the promise and good performance of the proposed methods ABC, ABC-CF, and ABC-ML.

6.2 Future works

The proposed ABC as a color feature was developed to solve the problem of image matching in clutter scenes. The ABC provided a new method to use color information for the robustness and effectiveness of image matching. In future work, we intend to expand the application of ABC to other subject areas. Meanwhile, absent colors can also be used in deep learning as a statistical feature of object classification.

Moreover, ABC, ABC-CF, and ABC-ML should be integrated into other fields for future work, such as object classification, defect detection, and pedestrian tracking.

Object classification

Object classification problems aim to identify targets with related characteristics while classifying the targets into two categories indicating positive and negative classes. The key point in classification is selecting features to analyze. In the present work, ABC was used to perform statistical classification and exhibited the advantages of robustness and distinguishability. Furthermore, color information may be input to the classifier to enable the model to learn the absent color features to perform object classification. We will extend ABC to this aspect in future work.

Defect detection

Defect detection is an important technique for product quality control in manufacturing industries. With the further development of production industries, defect detection may expend to consider an increasing set of problems, such as minor color printing errors. The ABC approach can perform defect detection by extracting defects as absent colors. It can also control the changes in illumination to a certain extent and does not require training a large number of templates to learn the features of defect-free data. To perform defect detection, a defect-free template can be divided into blocks, and then the corresponding blocks compared to search for absent colors. Therefore, we would like to apply the ABC approach to defect detection in future work.

Pedestrian tracking

Pedestrian tracking has been applied widely in video surveillance, human-computer interaction, and unmanned driving. It is a vital research topic in computer vision. Although many challenges remain in the process of pedestrian tracking, such as rotation,

deformation, and scale variation, ABC is robust to these as a color histogram-based approach. It differs from other approaches based on color histograms and improves on the sensitivity of other approaches to minor but significant colors. Furthermore, ABC has exhibited relatively good discrimination in matching or tracking similar people in pedestrian tracking applications. In future work, we will consider a detailed design to integrate ABC or combine it with our proposed ABC-CF and ABC-ML for pedestrian tracking applications.

Acknowledgements

I would like to express my sincere appreciation to my supervisor Prof. Takayuki Tanaka, who provided a great deal of assistance and valuable comments on our research. Your wise guidance has always helped me think logically and seek a deeper understanding. Thank you for your invaluable help during the 2^{nd} and 3^{rd} Ph.D grades. Moreover, I thank you from the bottom of my heart for your warmth and kindness over the years. I hope that I will be able to successfully complete the submission of this Ph.D. thesis with your continued assistance. Thank you again.

Thank you very much, Prof. Mineichi Kudo and Prof. Satoshi Kanai, for reviewing my thesis and providing so many key suggestions to finish my thesis and prepare the presentation.

Thank you, Prof. Shun'ichi Kaneko, my supervisor, for giving me the opportunity to study in Japan. In 2018, I came to Japan to start my Ph.D. study. You have guided me patiently, like a family member. You taught me the philosophy of learning and helped me understand how to think scientifically. You taught me how to express my research through presentation simply and vividly, as if telling a story. You also gave me a deeper understanding and cognition of image processing. At LM and GM, you always make vital suggestions while guiding me on how to dig into the essence of the problem. I am honored to be your last Ph.D. student. You have given me a great deal of tolerance and encouragement to keep me moving forward. Thanks, Prof. Shun'ichi Kaneko, all for this wonderful and unforgettable study life with you. This is a very important three and a half years in my life.

Thanks to Mr. So Sasatani, Dr. Masaya Itoh, and Dr. Yuan Li at Hitachi Ltd. for their suggestions and discussions given me important tips to improve my research. Thanks to Prof. Ming Fang from Changchun University of Science and Technology, China, for recommending me to study in Japan and guiding my study. Thank you very much to Ms. Miki Kikkawa for her support in both my research and my daily life.

I would like to thank all the staffs and members in HCE laboratory. Thanks to Dr. Wenjun Zhou for picking me up at the airport when I first came to Japan. Thanks to Dr. Sheng Xiang for helping me analyze the problems encountered in my research and share the solutions. Thanks to Dr. Yaping Yan for giving me a lot of sample materials. Thanks to Dr. Asahi Matsuda, Mr. Michihiro Yoshida, Mr. Kanta Matsuyama, and Ms. Akari Oshima. They considerably help me in my daily life. Thank you my friends, Mr. Hongyuan Ren and Mr. Guodong Wei; I learned a great deal from you.

Thank you very much to my family as well. It was their unconditional support and

encouragement that allowed me to successfully complete my Ph.D. study. Thank you, parents, for sharing your life with me in the weekly video. Thank you to my little brother for always mentioning me in his composition. I love them.

Finally, I appreciate the Ministry of Education, Culture, Sports, Science, and Technology Japan and Hokkaido University, who provide me abundant financial support.

References

[1] X. Zenggang, T. Zhiwen, C. Xiaowen, Z. Xue-min, Z. Kaibin, and Y. Conghuan, "Research on image retrieval algorithm based on combination of color and shape features," *Journal of signal processing systems*, vol. 93, no. 2, pp. 139–146, 2021.

[2] S. Zhang, Y. Wu, and J. Chang, "Survey of image recognition algorithms," in *2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, vol. 1. IEEE, 2020, pp. 542–548.

[3] A. Brunetti, D. Buongiorno, G. F. Trotta, and V. Bevilacqua, "Computer vision and deep learning techniques for pedestrian detection and tracking: A survey," *Neurocomputing*, vol. 300, pp. 17–33, 2018.

[4] T. Bo, K. Jianyi, and W. Shiqian, "Review of surface defect detection based on machine vision," *Journal of Image and Graphics*, vol. 22, no. 12, pp. 1640–1663, 2017.

[5] J. Ma, X. Jiang, A. Fan, J. Jiang, and J. Yan, "Image matching from handcrafted to deep features: A survey," *International Journal of Computer Vision*, vol. 129, no. 1, pp. 23–79, 2021.

[6] Z. Zivkovic and B. Krose, "An em-like algorithm for color-histogram-based object tracking," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, vol. 1. IEEE, 2004, pp. I–I.

[7] J. Ning, L. Zhang, D. Zhang, and C. Wu, "Robust object tracking using joint color-texture histogram," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 23, no. 07, pp. 1245–1263, 2009.

[8] M. J. Swain and D. H. Ballard, "Indexing via color histograms," in *Active Perception and Robot Vision*. Springer, 1992, pp. 261–273.

[9] M. A. Stricker and M. Orengo, "Similarity of color images," in *Storage and retrieval for image and video databases III*, vol. 2420. International Society for Optics and Photonics, 1995, pp. 381–392.

[10] P. Chang and J. Krumm, "Object recognition with color cooccurrence histograms," in *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, vol. 2. IEEE, 1999, pp. 498–504.

[11] J. Han and K.-K. Ma, "Fuzzy color histogram and its use in color image retrieval," *IEEE Transactions on image Processing*, vol. 11, no. 8, pp. 944–952, 2002.

[12] H. Shao, Y. Wu, W. Cui, and J. Zhang, "Image retrieval based on mpeg-7 dominant color descriptor," in *2008 The 9th International Conference for Young Computer Scientists*. IEEE, 2008, pp. 753–757.

[13] N.-C. Yang, W.-H. Chang, C.-M. Kuo, and T.-H. Li, "A fast mpeg-7 dominant color extraction with new similarity measure for image retrieval," *Journal of visual communication and image representation*, vol. 19, no. 2, pp. 92–105, 2008.

[14] J. R. Smith and S.-F. Chang, "Single color extraction and image query," in *Proceedings., International Conference on Image Processing*, vol. 3. IEEE, 1995, pp. 528–531.

[15] J. Huang, S. R. Kumar, M. Mitra, W.-J. Zhu, and R. Zabih, "Image indexing using color correlograms," in *Proceedings of IEEE computer society conference on Computer Vision and Pattern Recognition*. IEEE, 1997, pp. 762–768.

[16] G. Pass, R. Zabih, and J. Miller, "Comparing images using color coherence vectors," in *Proceedings of the fourth ACM international conference on Multimedia*, 1997, pp. 65–73.

[17] G. Qiu, "Color image indexing using btc," *IEEE transactions on image processing*, vol. 12, no. 1, pp. 93–101, 2003.

[18] S. Soleimanizadeh, D. Mohamad, T. Saba, and A. Rehman, "Recognition of partially occluded objects based on the three different color spaces (rgb, ycbcr, hsv)," *3D Research*, vol. 6, no. 3, pp. 1–10, 2015.

[19] A. Kadir, L. E. Nugroho, A. Susanto, and P. I. Santosa, "Foliage plant retrieval using polar fourier transform, color moments and vein features," *arXiv preprint arXiv:1110.1513*, 2011.

[20] T. Saikrishna, A. Yesubabu, A. Anandarao, and T. S. Rani, "A novel image retrieval method using segmentation and color moments," *Advanced Computing*, vol. 3, no. 1, p. 75, 2012.

[21] M. Mosbah and B. Boucheham, "The influence of the color model on the performance of a cbir system based on color moments," *Journal of Communication and Computer*, vol. 11, no. 3, pp. 266–273, 2014.

[22] P. A. Mlsna and J. J. Rodríguez, "Efficient indexing of multi-color sets for content-based image retrieval," in *4th IEEE Southwest Symposium on Image Analysis and Interpretation*. IEEE, 2000, pp. 116–120.

[23] R. Ravani, M. R. Mirali, and M. Baniasadi, "Parallel cbir system based on color coherence vector," in *17th International Conference on Systems, Signals and Image Processing*, 2010.

[24] A. Al-Hamami and H. Al-Rashdan, "Improving the effectiveness of the color coherence vector." *Int. Arab J. Inf. Technol.*, vol. 7, no. 3, pp. 324–332, 2010.

[25] J. Singh, A. Bajaj, A. Mittal, A. Khanna, and R. Karwayun, "Content based image retrieval using gabor filters and color coherence vector," in *2018 IEEE 8th International Advance Computing Conference (IACC)*. IEEE, 2018, pp. 290–295.

[26] Q. Zhao and H. Tao, "Object tracking using color correlogram," in *2005 IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*. IEEE, 2005, pp. 263–270.

[27] W. Rasheed, Y. An, S. Pan, I. Jeong, J. Park, and J. Kang, "Image retrieval using maximum frequency of local histogram based color correlogram," in *2008 International Conference on Multimedia and Ubiquitous Engineering (mue 2008)*. IEEE, 2008, pp. 62–66.

[28] V. Vinayak and S. Jindal, "Cbir system using color moment and color autocorrelogram with block truncation coding," *International Journal of Computer Applications*, vol. 161, no. 9, pp. 1–7, 2017.

[29] C. Palm, "Color texture classification by integrative co-occurrence matrices," *Pattern recognition*, vol. 37, no. 5, pp. 965–976, 2004.

[30] A. Vadivel, S. Sural, and A. K. Majumdar, "An integrated color and intensity co-occurrence matrix," *Pattern recognition letters*, vol. 28, no. 8, pp. 974–983, 2007.

[31] C.-H. Lin, R.-T. Chen, and Y.-K. Chan, "A smart content-based image retrieval system based on color and texture feature," *Image and Vision Computing*, vol. 27, no. 6, pp. 658–665, 2009.

[32] O. Losson, A. Porebski, N. Vandenbroucke, and L. Macaire, "Color texture analysis using cfa chromatic co-occurrence matrices," *Computer Vision and Image Understanding*, vol. 117, no. 7, pp. 747–763, 2013.

[33] J. Thangarasu and P. Geetha, "Content based image retrieval using quad tree block truncation coding with color co-occurrence feature for the big data platform," *Journal of Computational and Theoretical Nanoscience*, vol. 14, no. 8, pp. 3874–3886, 2017.

[34] M. Leila, "Registration techniques for multisensor remotely sensed imagery," *Photogrammetric Engineering and Remote Sensing*, vol. 62, no. 9, pp. 1049–1056, 1996.

[35] B. Zitova and J. Flusser, "Image registration methods: a survey," *Image and vision computing*, vol. 21, no. 11, pp. 977–1000, 2003.

[36] J. Ma, J. C.-W. Chan, and F. Canters, "Fully automatic subpixel image registration of multiangle chris/proba data," *IEEE transactions on geoscience and remote sensing*, vol. 48, no. 7, pp. 2829–2839, 2010.

[37] H.-M. Chen, M. K. Arora, and P. K. Varshney, "Mutual information-based image registration for remote sensing data," *International Journal of Remote Sensing*, vol. 24, no. 18, pp. 3701–3706, 2003.

[38] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *IEEE Transactions on systems, man, and cybernetics*, no. 6, pp. 610–621, 1973.

[39] H. Tamura, S. Mori, and T. Yamawaki, "Textural features corresponding to visual perception," *IEEE Transactions on Systems, man, and cybernetics*, vol. 8, no. 6, pp. 460–473, 1978.

[40] A. Laine and J. Fan, "Texture classification by wavelet packet signatures," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 15, no. 11, pp. 1186–1191, 1993.

[41] D. Dunn, W. E. Higgins, and J. Wakeley, "Texture segmentation using 2-d gabor elementary functions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 2, pp. 130–149, 1994.

[42] D. Zhang, M. M. Islam, G. Lu, and I. J. Sumana, "Rotation invariant curvelet features for region based image retrieval," *International journal of computer vision*, vol. 98, no. 2, pp. 187–201, 2012.

[43] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[44] T. Jabid, M. H. Kabir, and O. Chae, "Local directional pattern (ldp)–a robust image descriptor for object recognition," in *2010 7th IEEE international conference on advanced video and signal based surveillance*. IEEE, 2010, pp. 482–487.

[45] S. Fadaei, R. Amirfattahi, and M. R. Ahmadzadeh, "Local derivative radial patterns: A new texture descriptor for content-based image retrieval," *Signal Processing*, vol. 137, pp. 274–286, 2017.

[46] X. Yang, S. Koknar-Tezel, and L. J. Latecki, "Locally constrained diffusion process on locally densified distance spaces with applications to shape retrieval," in *2009 IEEE conference on computer vision and pattern recognition*. IEEE, 2009, pp. 357–364.

[47] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, vol. 1. Ieee, 2005, pp. 886–893.

[48] D. K. Park, Y. S. Jeon, and C. S. Won, "Efficient use of local edge histogram descriptor," in *Proceedings of the 2000 ACM workshops on Multimedia*, 2000, pp. 51–54.

[49] V. Torre and T. Poggio, "On image analysis by methods of moments," *Journal of the Optical Society of America*, vol. 70, pp. 920–930, 1980.

[50] J. Iivarinen, M. Peura, J. Särelä, and A. Visa, "Comparison of combined shape descriptors for irregular objects." in *BMVC*. Citeseer, 1997.

[51] M. J. Swain and D. H. Ballard, "Color indexing," *International journal of computer vision*, vol. 7, no. 1, pp. 11–32, 1991.

[52] N. K. Verma, A. Goyal, A. Chaman, R. K. Sevakula, and A. Salour, "Template matching for inventory management using fuzzy color histogram and spatial filters," in *2015 IEEE 10th Conference on Industrial Electronics and Applications (ICIEA)*. IEEE, 2015, pp. 317–322.

[53] B. Singh and B. Mazumdar, "Content retrieval from xray images using color & texture features," *Methodology*, vol. 1, p. 6, 2010.

[54] K. P. Jasmine and P. R. Kumar, "Integration of hsv color histogram and lmebp joint histogram for multimedia image retrieval," in *Intelligent Computing, Networking, and Informatics*. Springer, 2014, pp. 753–762.

[55] S. Zeng, J. Bai, and R. Huang, "Image retrieval based on color-spatial histograms," in *The 2014 2nd International Conference on Systems and Informatics (ICSAI 2014)*. IEEE, 2014, pp. 780–784.

[56] A. Gupta and M. Gangadharappa, "Image retrieval based on color, shape and texture," in *2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)*. IEEE, 2015, pp. 2097–2104.

[57] S. Lijuan and H. Fengqi, "Research on colour and texture feature based image retrieval," in *2015 International Conference on Intelligent Transportation, Big Data and Smart City*. IEEE, 2015, pp. 626–628.

[58] X.-m. Shen, J.-x. Zhou, and T.-W. Xu, "Minority costume image retrieval by fusion of color histogram and edge orientation histogram," in *2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)*. IEEE, 2016, pp. 1–7.

[59] D. Soni and K. Mathai, "An efficient content based image retrieval system based on color space approach using color histogram and color correlogram," in *2015 Fifth International Conference on Communication Systems and Network Technologies*. IEEE, 2015, pp. 488–492.

[60] H. Lacheheb and S. Aouat, "Simir: New mean sift color multi-clustering image retrieval." *Multimedia Tools & Applications*, vol. 76, no. 5, 2017.

[61] M. M. Elsheh and S. A. Eltomi, "Content based image retrieval using color histogram and discrete cosine transform," *International Journal of Computer Trends and Technology*, vol. 19, no. 9, pp. 25–31, 2019.

[62] A. K. Bhunia, A. Bhattacharyya, P. Banerjee, P. P. Roy, and S. Murala, "A novel feature descriptor for image retrieval by combining modified color histogram and diagonally symmetric co-occurrence texture pattern," *Pattern Analysis and Applications*, vol. 23, no. 2, pp. 703–723, 2020.

[63] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 25, no. 5, pp. 564–577, 2003.

[64] L. Bertinetto, J. Valmadre, S. Golodetz, O. Miksik, and P. H. Torr, "Staple: Complementary learners for real-time tracking," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1401–1409.

[65] H. Zhang, G. Liu, and Z. Hao, "Robust visual tracking via multi-feature response maps fusion using a collaborative local-global layer visual model," *Journal of Visual Communication and Image Representation*, vol. 56, pp. 1–14, 2018.

[66] H. A. Abdelali, F. Essannouni, L. Essannouni, and D. Aboutajdine, "Fast and robust object tracking via accept–reject color histogram-based method," *Journal of Visual Communication and Image Representation*, vol. 34, pp. 219–229, 2016.

[67] H. Possegger, T. Mauthner, and H. Bischof, "In defense of color-based model-free tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 2113–2120.

[68] A. Lukezic, T. Vojir, L. ˇCehovin Zajc, J. Matas, and M. Kristan, "Discriminative correlation filter with channel and spatial reliability," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 6309–6318.

[69] J. Fan, H. Song, K. Zhang, Q. Liu, and W. Lian, "Complementary tracking via dual color clustering and spatio-temporal regularized correlation learning," *IEEE Access*, vol. 6, pp. 56 526–56 538, 2018.

[70] Z. Hao, G. Liu, J. Gao, and H. Zhang, "Robust visual tracking using structural patch response map fusion based on complementary correlation filter and color histogram," *Sensors*, vol. 19, no. 19, p. 4178, 2019.

[71] S. P. Mudunuri and S. Biswas, "Low resolution face recognition across variations in pose and illumination," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 5, pp. 1034–1040, 2015.

[72] B. Liu, X. Shu, and X. Wu, "Fast screening algorithm for rotation invariant template matching," in *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2018, pp. 3708–3712.

[73] H. Lee, H. Kwon, R. M. Robinson, and W. D. Nothwang, "Dtm: Deformable template matching," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 1966–1970.

[74] X. Lan, X. Zhu, and S. Gong, "Person search by multi-scale matching," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 536–552.

[75] N. McLaughlin, J. Ming, and D. Crookes, "Largest matching areas for illumination and occlusion robust face recognition," *IEEE transactions on cybernetics*, vol. 47, no. 3, pp. 796–808, 2016.

[76] A. Kaur and B. Kranthi, "Comparison between ycbcr color space and cielab color space for skin color segmentation," *International Journal of Applied Information Systems*, vol. 3, no. 4, pp. 30–33, 2012.

[77] Y. Tian, S. Kaneko, S. Sasatani, M. Itoh, and M. Fang, "Robust and precise matching algorithm combining absent color indexing and correlation filter," *Information*, vol. 12, no. 10, p. 428, 2021.

[78] L. Wilkinson, "Revising the pareto chart," *The American Statistician*, vol. 60, no. 4, pp. 332–334, 2006.

[79] S. Lee, J. H. Xin, and S. Westland, "Evaluation of image similarity by histogram intersection," *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*, vol. 30, no. 4, pp. 265–274, 2005.

[80] K. Grauman and T. Darrell, "The pyramid match kernel: Efficient learning with sets of features," *Journal of Machine Learning Research*, vol. 8, no. Apr, pp. 725–760, 2007.

[81] H. Chen, K. Xie, H. Wang, and C. Zhao, "Scene image classification using locality-constrained linear coding based on histogram intersection," *Multimedia Tools and Applications*, vol. 77, no. 3, pp. 4081–4092, 2018.

[82] N. H. Anderson, "Note on weighted sum and linear operator models," *Psychonomic Science*, vol. 1, no. 1, pp. 189–190, 1964.

[83] B. Fuglede and F. Topsoe, "Jensen-shannon divergence and hilbert space embedding," in *Proceedings of the International Symposium onInformation Theory.* ISIT, June 2004.

[84] B. Bigi, "Using kullback-leibler distance for text categorization," in *European Conference on Information Retrieval.* ECIR, 2003, pp. 305–319.

[85] A. Bhattacharyya, "On a measure of divergence between two statistical populations defined by their probability distributions," *Bull. Calcutta Math. Soc.*, vol. 35, pp. 99–109, 1943.

[86] Y. Wu, J. Lim, and M.-H. Yang, "Online object tracking: A benchmark," in *Proceedings of the IEEE conference on computer vision and pattern recognition.* CVPR, 2013, pp. 2411–2418.

[87] M. Wang, Y. Liu, and Z. Huang, "Large margin object tracking with circulant feature maps," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4021–4029.

[88] J. Ning, J. Yang, S. Jiang, L. Zhang, and M.-H. Yang, "Object tracking via dual linear structured svm and explicit feature map," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 4266–4274.

[89] L. Bertinetto, J. Valmadre, J. F. Henriques, A. Vedaldi, and P. H. Torr, "Fully-convolutional siamese networks for object tracking," in *European conference on computer vision.* ECCV, 2016, pp. 850–865.

[90] J. Cheng, Y.-H. Tsai, W.-C. Hung, S. Wang, and M.-H. Yang, "Fast and accurate online video object segmentation via tracking parts," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* CVPR, 2018, pp. 7415–7424.

[91] Y. Tian, S. Kaneko, S. Sasatani, and M. Itoh, "Robust picture search by absent color indexing," in *Proceedings of the Seventh Asia International Symposium on Mechatronics.* Springer Singapore, 2020, pp. 860–866.

[92] Y. Tian, S. Kaneko, S. Sasatani, M. Itoh, and M. Fang, "Reliable and accurate pattern search by combination of absent color indexing with correlation filter," in *Proceedings of the IECON 2019 Forty-fifth Annual Conference of the IEEE Industrial Electronics Society.* IEEE, 2019, pp. 5273–5278.

[93] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, "Visual object tracking using adaptive correlation filters," in *2010 IEEE computer society conference on computer vision and pattern recognition*. IEEE, 2010, pp. 2544–2550.

[94] Y. Li and J. Zhu, "A scale adaptive kernel correlation filter tracker with feature integration," in *European conference on computer vision*. Springer, 2014, pp. 254–265.

[95] A. Mahalanobis, B. V. Kumar, S. Song, S. R. F. Sims, and J. F. Epperson, "Unconstrained correlation filters," *Applied Optics*, vol. 33, no. 17, pp. 3751–3759, 1994.

[96] H. Grad, *Proceedings of Symposia in Applied Mathematics*. American Mathematical Society Providence, Rhode Island, 1967.

[97] M. Danelljan, G. Häger, F. Khan, and M. Felsberg, "Accurate scale estimation for robust visual tracking," in *British Machine Vision Conference, Nottingham, September 1-5, 2014*. BMVA Press, 2014.

Appendix A Publications lists

A.1 Journal Paper

1. Tian Y, Kaneko S, Sasatani S, Itoh M, Fang M. Robust and Precise Matching Algorithm Combining Absent Color Indexing and Correlation Filter. Information. 2021; 12(10):428. https://doi.org/10.3390/info12100428

A.2 International Conferences

1. Tian, Y., Kaneko, S. I., Sasatani, S., Itoh, M., & Fang, M. Reliable and Accurate Pattern Search by Combination of Absent Color Indexing with Correlation Filter. In IECON 2019-45th Annual Conference of the IEEE Industrial Electronics Society, Vol. 1, pp. 5413-5418. DOI: 10.1109/IECON.2019.8927524.

2. Tian, Y., Kaneko, S. I., Sasatani, S., & Itoh, M. Robust Picture Search by Absent Color Indexing. In Proceedings of the Seventh Asia International Symposium on Mechatronics, Springer, Singapore, 2020, Vol. 589, pp. 860-866. https://doi.org/10.1007/978-981-32-9441-7_88

A.3 Domestic Conferences

1. Tian, Y., Kaneko, S. I., Oshima, A., Sasatani, S., and Itoh, M. Absent Color Indexing and Its Applications to Robust Pattern Search. The 25th Symposium on Sensing via Image Information. 2019.

2. Tian, Y., Kaneko, S. I., Sasatani, S., Itoh, M., Fang, M., and Tanaka, T. Absent Color Indexing and Its Applications to store vision. The 26th Symposium on Sensing via Image Information. 2020.

3. Tian, Y., Kaneko, S. I., Sasatani, S., Itoh, M., Fang, M., and Tanaka, T. Absent Color Indexing under Aggregation with Minor Colors for Robust Pattern Search. Dynamic Image processing for real Application workshop 2020. 2020.