Identification and Extraction of Information from Prescription Pill Bottles

Group 13: Cameron Chong, Emily Earl, Rachelle Gelden,
Jasper Looman, Brandon Unger

March 2019

Abstract

Many individuals, particularly those in the elderly community, struggle with the size of the font written on medication bottles. This can result in individuals taking their medication improperly, causing undesired side effects or limiting the drugs functionality. As a means of mitigating this issue, this paper outlines an algorithm used to extract the information written on prescription medication bottles and portray the extracted text in a manner such that the font size can be made larger for easier viewing.

Contents

1 Introduction and Background

As the science revolving around medications continues to evolve and expand, continually more and more conditions will be provided with treatments. While these advancements in medicine are providing solutions to a large amount of problems that exist within society, if these new treatment plans are not being followed as prescribed the value that society will gain from them reduces substantially. A study was performed that showed that 33% of all errors regarding proper medication consumption can be attributed to confusion brought about by the packaging or labelling of medications (Jeetu Girish, 2010).

A particular group of people who require the benefits from these treatments the most, the elderly, seem to have the largest number of issues regarding medication bottles and the correct dosing of prescriptions. Based on an experimental group of 59 individuals over the age of 70, Notenboom et al. discovered that over 20% of the individuals thought that the text on the bottle was too small to read effectively (Notenboom, 2014). While this may not be the largest issue that the elderly face with the design of medication packaging, it is one of many that contributes to the potential clinical consequences that can arise due to misuse of prescribed drugs.

In addition to the elderly, it has been found that the minimum font size that should be used to encapsulate the majority of society should be no less than 9 points, and no less than 16 for the visually impaired, however, the bulk of prescription bottles list the medication instructions at a font of point 9.3, barely over the minimum for those with perfect vision (Text Matters, 2001). This suggests that in addition to the elderly, a large number of people are likely struggling with identifying the text listed on medication bottles.

It has also been found that those with movement disorders, such as Parkinson's or Essential Tremor, have increased difficulty reading small fonts due to axial rigidity (Parrales Bravo et al., 2017).

As a means of avoiding the clinical issues that result from incorrect dosages of medications that occur within the elderly community as well as the visually impaired due to the extremely small print used on medication bottles, a tool should be developed to extract the information from pill bottles and make it more easily accessible to the general user.

2 Related Work

While not using computer vision, a variety of companies have attempted to alleviate this issue by removing the need for labels entirely - developing "smart" bottles that are are programmed by a caregiver or medication prescriber to apply an auditory and/or visual reminder for when a certain drug should be taken ("AdhereTech — Home," n.d.). Other technologies are available for the visually impaired such as Talking RX, which allows the medication prescriber to record a message on a small device attached to the bottom of the bottle which can be re-played more than 100 times ("Talking Rx - Product Profile - American Foundation for the Blind," n.d.).

Many larger companies, such as Google and Microsoft, have developed computer vision tools that can be used to extract small amounts of text, such as the camera feature on the Google Assistant App, as well as the Microsoft Seeing AI app ("How To Make Medication Bottles Accessible For Vision Impairment," 2018). These tools can extract text, display it larger on one's device, and even output it audibly. However, as the cylindrical nature of pill bottles usually creates a situation where the whole label cannot be seen in a single frame, they can produce skewed results and can cut off key parts of information when reading the label to the user.

This leads to the necessity of a way to combine information gained from multiple images to allow for the full label to be accurately depicted. This can be accomplished by image stitching, a process used to apply appropriate overlap between an array of images, a process commonly known for its implementation in the development of panoramic images. A certain procedure developed by Levin et al. (2004) performs image stitching by using the gradient domain, allowing the disturbed edges along the image seam to be minimized.

In addition, as the cylindrical nature of the bottle causes some of the text to be distorted, a method needs to be implemented to effectively unwrap the cylindrical depiction of the label gained from the images and convert it to a rectangular form for the image stitching process. A cylinder unwrapping method was developed by Zosel (2015) to determine data from a series of rock cores, however, the general functionality of the method can be extended to a variety of cylindrical objects.

3 Proposed Solution

Our proposed implementation was based off of concepts from both the works of Zosel (2015) and Levin et al. (2004) to produce an algorithm that can unwrap images of cylindrical surfaces and stitch the unwrapped images together. The algorithm was divided into four main steps: reading the input images, object identification and image correction, flattening the detected label, and stitching the images together. These four main steps were completed based on the assumptions, as follows:

1. The set of images provided by the user must collectively give a complete view of the label on the bottle.

2. Photos must be taken on a horizontal and level surface with a dark back-

ground.

3. Photos must all be taken at the same angle which must be faced straight on to the center of the pill bottle.
4. The bottle's label must approximately be the center of the image.
5. The label must be white with black text, and the label must be brighter than the bottle.
6. The height of the bottle is greater than the diameter of the bottle.
7. The bottle is only being rotated in the z-direction (see Fig. 1).
8. The label is the largest connected component in the image at the center of the image.
9. Camera distortion is considered to be negligible.
10. The width of the label is equal to width of the bottle.

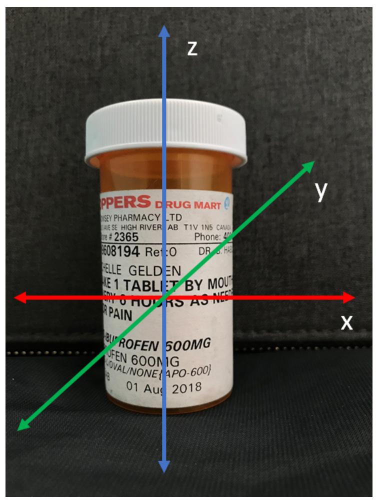


Figure 1: Bottle Orientation

The first step of our algorithm is reading in a set of images of a prescription bottle from the user. The set of images is captured by rotating the pill bottle around the z-axis (see Fig. 1). This process must last one full rotation of the bottle. The photos must all be taken at the same angle and the photos must be taken straight on relative to the bottle. The bottle must also be approximately in the center of each image.

The second step of our algorithm is label identification and image correction. This step examines each image and locates the label within the image. This is achieved by finding the connected component with the largest area whose bounding box encapsulates the center of the image (by Assumption 4). Assumption 2 is needed as it ensures that the label is detected as the largest connected component in the image. Using only the images in which labels have been detected, a new set of input images is created by cropping the labels out of the photos. The algorithm uses morphological processing and thresholding to detect the bounding box of the label. It is important to note that we assume the center of this label is in-line with the center of the cylinder and the camera. The orientation of each label is then corrected by using eigenvectors to identify the major and minor axes. Next, we calculate the average x and y sizes of all labels. We then scale each detected label to the previously calculated average x and y sizes.

The third step of our algorithm flattens each detected label from the 2D projection of a 3D object (image of the label) to a 2D projection (label as a plane). By Assumption 10, we can calculate the radius of the bottle as half the width of the label. We assume our origin 0 to be the center of the label as seen in Figure 2.A. We define our boundaries x to be the edges of the label. We define our output image to be the same size as the input image.

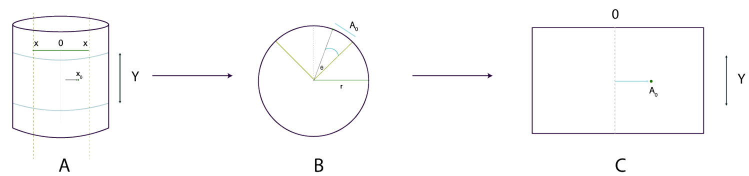


Figure 2: Image Unwrapping Process Diagram

For every column x_i inside the boundaries x, we use Equation (1) to calculate the approximate angle, θ_i, that column x_i exists on the bottle away from origin 0 as seen in Figure 2.B.

$$\theta_i = sin^{-1}(x_i/r) \tag{1}$$

We then calculate the arc length A_i of column x_i away from origin 0 by Equation (2). This step uses uses the center and the width of the detected label to estimate a transformation that flattens the curved texture to undo the distortion caused by the cylinder. We used the reverse of this transformation in order to get a complete mapping of intensities for the output image. Since the pixels on the outer edges of the bottle are more distorted and stretched out than the pixels in the centre, we then take the value of the pixels in the column of the original image and place them in our output image at a distance of A_i from the origin as seen in Figure 2.C.

$$A_i = r\theta_i \tag{2}$$

We then take the value of the pixels in the column of the original image and place them in our output image at a distance of A_i from the origin as seen in Figure 2.C.

Since we are working in a non-continuous domain, the mapping from x_i to A_i does not span A_i, therefore, we cannot use a forward mapping. Thus, it is required that our calculations above be completed in reverse, resulting in the final equation as shown in Equation 3. Because we are working in a non-continuous domain the mapping from x_i to A_i does not span A_i, therefore, we cannot use a forward mapping. Thus, it is required that our calculations above be completed in reverse, resulting in the final equation as shown in Equation 3.

$$x_i = sin(A_i/r) * r \tag{3}$$

The fourth step of our algorithm takes the flattened surface texture for each label and stitches them together. This stitching is done using the Histogram of Oriented Gradients (HoG) algorithm to find the best alignment of two consecutive images. This is done by comparing the HoG descriptor across the x-axis and finding the minimum weighted vector which is used to determine optimal image alignment. We used the HoG descriptor instead of pixel intensity to determine the optimal image alignment because slight offsets between images will have less effect on the output since the HoG algorithm uses groups of pixels as cells.

The final step involves using text detection on the flattened and stitched complete view of the label. We use MatLab's Computer Vision System Toolbox OCR Text Recognition module on the image and output detected text to a text file, textDetected.txt.

In conclusion, our algorithm takes a series of images of a pill bottle, applies an orientation correction to each image and detects the label in every images. Each label is then transformed and stitched together to form a complete view of the label. Text is then detected from this label and outputted to a text file.

4 Experimental Results

The first step of our algorithm involves label detection and image correction. The label is first identified, then it corrects the angle of the label so that it is vertical as shown in Figures 3 through 5.

This step fails when there are other objects in the picture that are either connected to the bottle such as hands holding the bottles or items in the background that are partially covered by the bottle (see Fig. 6 and Fig 7).

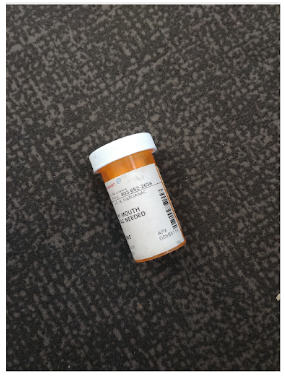


Figure 3: Crooked Image Given as Input

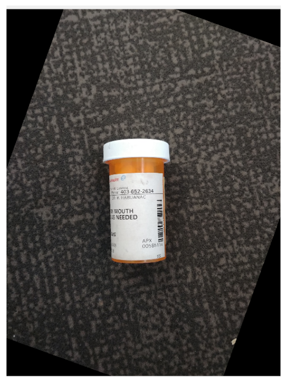


Figure 4: Bottle Orientation Corrected

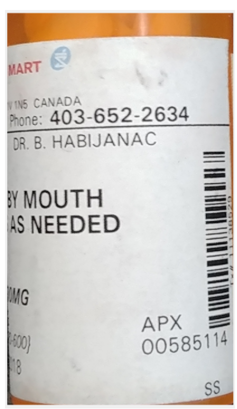


Figure 5: Label Identified and Given as Output

Figure 6: Invalid Input Image

Figure 7: Incorrect Label Detection in Output

The flattening step of our algorithm unwraps the label from the cylindrical pill bottle as shown in Fig. 8 and Fig. 9. This involves transforming a 2D projection of a 3D object to the original 2D existence. As seen in Fig 10. and Fig 11. if given invalid input, the label flattening algorithm cannot correctly identify the origin of the label and therefore cannot transform it correctly to flatten the label.

Figure 8: Image Given as Input

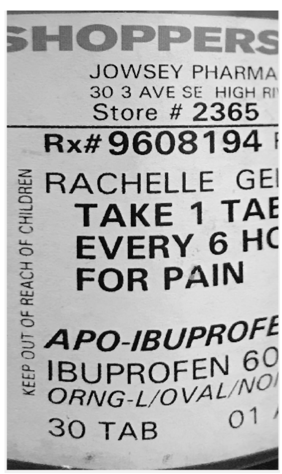


Figure 9: Flattened Image Given as Output

Figure 10: Invalid Input Image

Figure 11: Incorrect Flattened Image Given as Output

In the image stitching step, we take a series of flattened images of the label and stitch the images together to output a complete image of the label as shown in Figures 12 through 16. When it is not possible to stitch the a series of images together, the stitching algorithm returns -1 and our algorithm fails.

Figure 12: First Input Image

Figure 13: Second Input Image

Figure 14: Third Input Image

Figure 15: Output of Stitching the Three Images

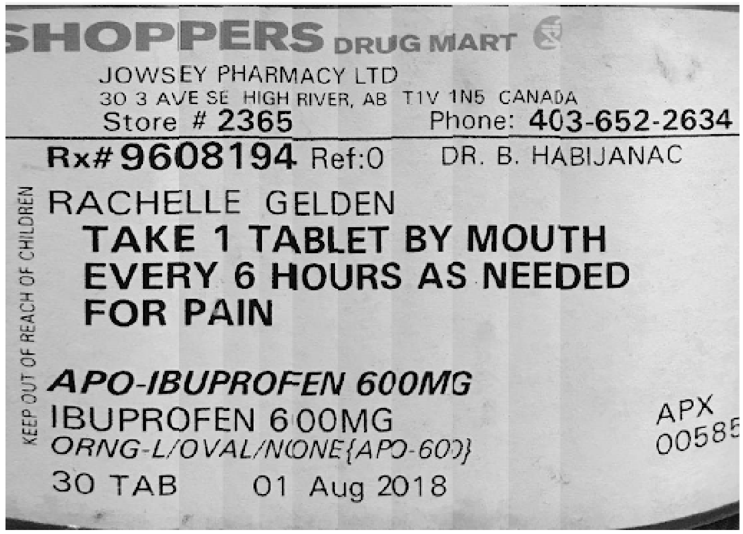


Figure 16: Output Image When Input Images Provide Complete View of Label

The text detection step of the algorithm takes the stitched image of the label and detects text on the label and writes the text to a text file. The text detection is not 100% accurate as it struggles in areas where the image was stitched or where remaining cylindrical and projection distortion is the most prominent. It also has difficulty determining the difference between different clusters of text when both horizontal and vertical text is given. An example of the text detection process is shown in Figures 17 and 18.

Text detection will not be successful if a prior step cannot be completed successfully. The accuracy of the text detection decreases as light, cylindrical and projection distortion become more prominent.

A visual representation of this process can be viewed at https://www.youtube.com /watch?v=W55owYO7Kmgfbclid=IwAR0xHvCgeIQQfTXXyPQ1Matn4KYAA mI0QnrcmjBKUGU_dGZDhsPtDIZsrxM

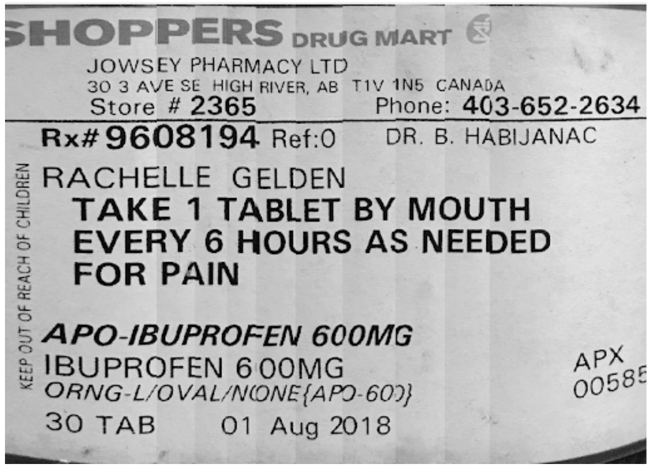
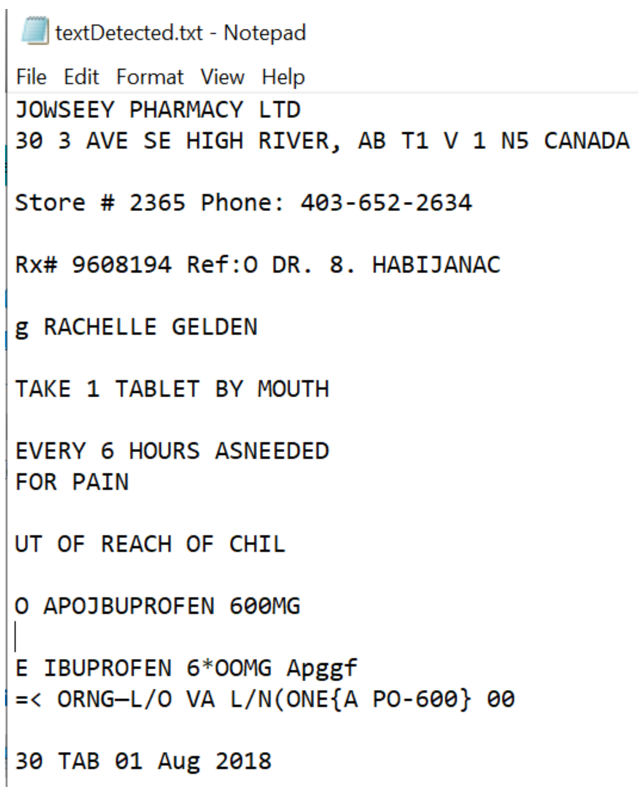


Figure 17: Input Image for Text Extraction


textDetected.txt - Notepad
File Edit Format View Help

JOWSEEY PHARMACY LTD
30 3 AVE SE HIGH RIVER, AB T1 V 1 N5 CANADA

Store # 2365 Phone: 403-652-2634

Rx# 9608194 Ref:O DR. 8. HABIJANAC

g RACHELLE GELDEN

TAKE 1 TABLET BY MOUTH

EVERY 6 HOURS ASNEEDED
FOR PAIN

UT OF REACH OF CHIL

O APOJBUPROFEN 600MG

E IBUPROFEN 6*OOMG Apggf
=< ORNG—L/O VA L/N(ONE{A PO-600} 00

30 TAB 01 Aug 2018

Figure 18: Output Text File

5 Discussion

In the object identification step of our algorithm, we achieve more accurate results when the pill bottle is photographed on a dark and opaque background as this limits reflection and glare in the image. Additionally, the algorithm's accuracy significantly decreases when there are other objects in the image that overlap or touch the pill bottle (ie. hands holding the bottle) as it can no longer detect only the label of the bottle. One strength of this portion of the algorithm is that it can detect a white label on the bottle without detecting the bottle's white lid as well. This is possible because by our assumptions, the label's connected component is always larger than the lid's connected component.

In the label flattening step of the algorithm, we achieve the best results when the bottle has a large radius and the picture of the bottle is taken from further away as this lessens the effect of the projection on the bottle. Additionally, our algorithm is much more accurate if the label spans most of the bottle in the image (Fig. 19 and Fig 20). This is because we can no longer estimate the location of the origin for our mapping function correctly which causes for an incorrect transformation in the flattening process. Inaccurate results are received when the picture is taken on an angle, or when the camera taking the image has too much distortion. Furthermore, we are unable to flatten the image if the bottle is has multiple label components (Fig. 21).

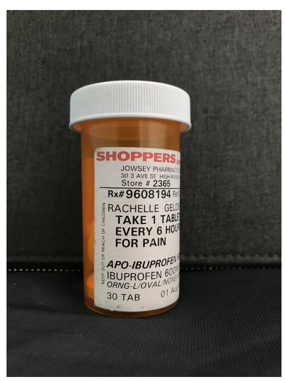


Figure 19: Undesirable Input Image

In the stitching step of our algorithm, we achieve the best results when all the images of the flattened labels have an approximately equal angle of alignment. This step works best when there are fewer images to stitch as our algorithm becomes linearly computationally more expensive with each input image added. Optimal results are produced when there is minimal curvature and distortion in the images to be stitched. Weaker results are achieved when the pictures of the bottle are taken too close as this causes a larger gradient shift between images and a larger projection distortion. Our stitching algorithm also fails

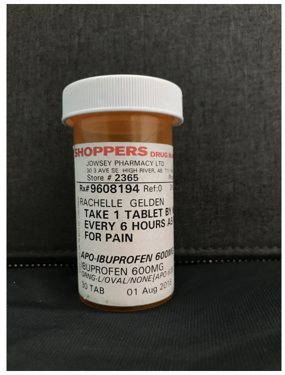


Figure 20: Desirable Input Image

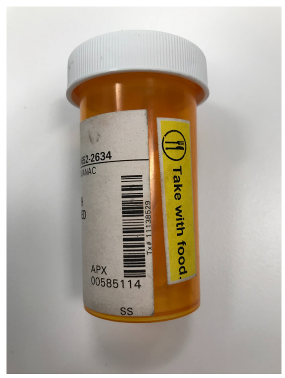


Figure 21: Bottle with Multiple Label Components

when trying to stitch any image that is on an angle or when the offset between images in the y-direction is greater then the height of the text on the bottle.

In the text detection step of our algorithm, the best results are achieved when the stitched flattened image generates clear and undistorted text. This step yields a close approximation of the text on the label, however it is rarely 100% accurate. The text detection step produces weaker results if the label has text in an unusual typeface or is handwritten ("Optical Character Recognition (OCR) Limitations - Meridian Outpost," n.d.). Additionally, if the image is distorted or there is text in both horizontal and vertical directions, the accuracy of detection decreases significantly.

Ultimately, while our algorithm is dependent on the outlined set of assumptions, its accuracy is primarily reliant on the quality of the input images provided by the user. Many improvements could be made to make the algorithm more robust, however the fundamental functionality for the algorithm is established and working.

6 Conclusion and Future Work

Although this algorithm did successfully combine multiple images of a pill bottle into coherent textual information, there are still many places where it could be improved. An example of this is that the runtime of the flattening algorithm could be improved by implementing vectorization. In addition, if provided time to alter the weighting algorithm of the image stitching process, the algorithm would not be limited to requiring the first and last images in the final result. In regards to the text detection step of the algorithmic process, a spell checking tool could be implemented to attempt and avoid simple mistakes and an algorithm could be developed to format the output to match a standardized form opposed to a outputting a long array of sentences. It would also be ideal to incorporate a method to be able to use a live video of someone rotating a bottle in their hand as the input for the system.

7 References

AdhereTech — Home. (n.d.). Retrieved March 31, 2019, from https://www.adhere tech.com/

How To Make Medication Bottles Accessible For Vision Impair ment. (2018, June 11). Retrieved March 31, 2019, from Veroniiiica website: https://veroniiiica .com/2018/06/11/how-to-make-medication-bottles-accessible-for-vision-impairment/

Jeetu, G., Girish, T. (2010). Prescription Drug Labeling Medication Errors: A Big Deal for Pharmacists. Journal of Young Pharmacists: JYP, 2(1), 107–111. https://doi.org/10.4103/0975-1483.62218

Levin, A., Zomet, A., Peleg, S., Weiss, Y. (2004). Seamless Image Stitching in the Gradient Domain. In T. Pajdla J. Matas (Eds.), Computer Vision - ECCV 2004(pp. 377–389). Springer Berlin Heidelberg.

Notenboom, K., Beers, E., van Riet-Nales, D. A., Egberts, T. C. G., Leufkens, H. G. M., Jansen, P. A. F., Bouvy, M. L. (2014). Practical Problems with Medication Use that Older People Experience: A Qualitative Study. Journal of the American Geriatrics Society, 62(12), 2339–2344. https://doi.org/10.1111/jgs.13126

Optical Character Recognition (OCR) Limitations - Meridian Outpost. (n.d.). Retrieved March 31, 2019, from http://www.meridianoutpost.com/resources/arti cles/ocr-limitations.php

Parrales Bravo, F., Del Barrio García, A. A., Gallego de la Sacristana, M., López Manzanares, L., Vivancos, J., Ayala Rodrigo, J. L. (2017). Support System to Improve Reading Activity in Parkinson's Disease and Essential Tremor Patients. Sensors (Basel, Switzerland), 17(5). https://doi.org/10.3390/s17051006

Talking Rx - Product Profile - American Foundation for the Blind. (n.d.). Retrieved March 31, 2019, from https://www.afb.org/prodProfile.asp?ProdID=799

Text Matters. (2001). typography for visually impaired people. Retrieved from http://www.textmatters.com/resources/pdfs/visImpd_typogTM.pdf

Zosel, Aaron. (2015). Development and Application of a Method for Unwrapping Single Images of Cylindrical Objects.

8 Appendix - Required Software

In order to run the our algorithms the Computer Vision System Toolkit, VLFeat and the OCR libraries must be installed. The Computer Vision System Toolkit is used for basic morphological, point and spatial processing when flattening the image. The VLFeat library is used to do HOG analysis on the image to determine the gradients of the images for image stitching. Finally, the OCR Library is used to analyze and detect the text from our flattened and stretched images.