

<p style="text-align: center;">Technology to help blind people : Image restitution by sonorous signals</p>

Abstract

This paper presents a project to render picture by auditory information. A picture is processed by computer to extract significant edges or, in other words, to detect the transition between two important areas. Besides this edges, the algorithm can provide a distance estimation following the level of blur in the picture when picture is focalized on near objects. The result of analyze is expressed in term of basic sounds. One aims of the project is to run the system on light and inexpensive hardware like mobile phone or pocket PC. The main purpose of the system is to give help to blind people in real time.

Author

First Name : Frederic

Last Name : Paquay

Phone : +32.(0)485.168.590

e-mail : paquay_frederic@yahoo.fr

Degree : University degree from the University of Liege (Belgium) obtained in 1995.

Background : 2 years of work in the University of Liege and then 7 years of work in EVS Broadcast Equipment, expert in development and image processing, currently workless.

<p style="text-align: center;">Technology to help blind people : Image restitution by sonorous signals</p>

Table of Contents

1 - Description.....	3
1.1 - Overview.....	3
1.2 - Video capture.....	3
1.3 - Picture analyze.....	4
1.3.1 - Preprocess.....	4
1.3.2 - Edge detection.....	5
1.3.3 - Distance estimation from blur level.....	5
1.4 - Sonorous rendering.....	7
1.4.1 - Rendering areas.....	8
1.4.2 - First approach.....	8
1.5 - Rendering parameters.....	9
2 - Project convenience.....	10
2.1 - Advantages.....	10
2.2 - Disadvantages.....	10
3 - Progress report.....	11
3.1 - Video capture.....	11
3.2 - Picture analysis.....	11
3.3 - Sonorous rendering.....	11
3.4 - Parameters control.....	12
4 - Prototype.....	13
5 - Conclusion.....	13
6 - Usefull links.....	14

Technology to help blind people : Image restitution by sonorous signals

1 - Description

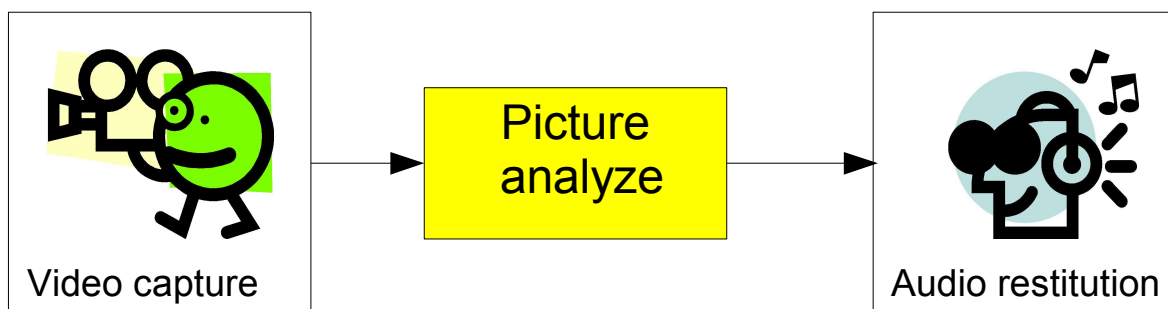
This project tries to develop a system to translate visual elements into auditory signals and thus to help blind people to have an idea about their visual surrounding. This system should be available for most people, its price should then be as low as possible. Despite this low price, the system should give a significant help to daily life of blind people. Unlike other systems, this one gives a large « vision » of the surrounding.

1.1 - Overview

The input of the system is a sequence of digital pictures from any kind of digital camera.

These pictures are then processed by an algorithm to extract significant edges and to compute an estimation of the distance from the camera. The distance can be computed with only one picture if the lens is focalized on nearest objects, the background is more and more blur as the object is far from the lens. The algorithm computes the level of blur to estimate the distance.

At the end of process, the edges' information with their distance is rendered as basic sonorous signals. Words could be used instead of basic sounds but this solution could be too difficult to implement for the expected result.



One important feature of the system is that blind people can select a lot of parameters so that he can feel better what he wants. Blind people have to do a big job to interpret information from the system.

1.2 - Video capture

The hardware to capture video could be any kind of digital camera as video camera, digital picture camera, smartphone camera, web cam,...

To get good capture, the camera has to respect as more as possible some features :

- a high resolution enough,

**Technology to help blind people :
Image restitution by sonorous signals**

- low picture noise,
- low capture time,
- high frequency of capture to get capture while walking,
- adaptative lens to very short depth of field, so that high level of blur can be obtained for far objects,
- low cost and light,
- easy connection to process system.

Moreover, for a much better confort, the video capture should be automatic (like video camera instead of picture camera) and it should be put on the head or near the head of the people, so that their hands remain free.

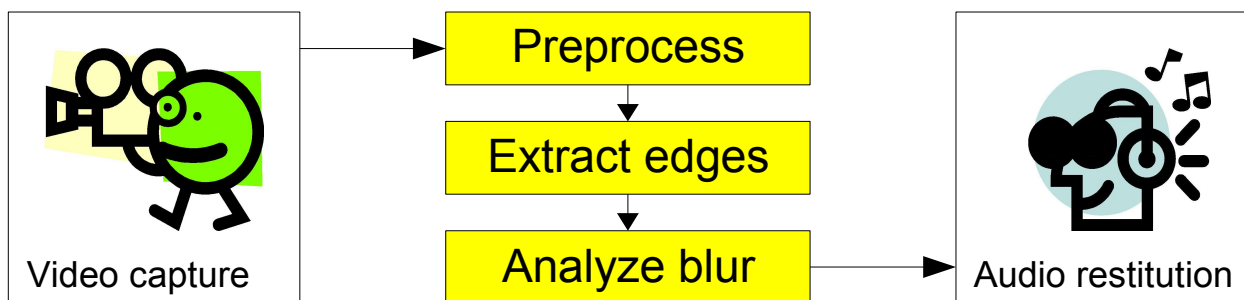
1.3 - Picture analyze

This step is the core of the process. From a picture, commonly a color picture with unfocalized areas (blur), the algorithm extracts significant edges with their approximate distance from camera following level of blur.

If this algorithm has to run on most hardware systems, even low performance systems like smartphone or pocket PC, it has to use low amount of process time and low amount of memory. Even if technology increases highly, memory size and high speed processor still have an important cost.

There is 3 steps in the whole process of the algorithm :

1. a picture preprocess to get fewer data to process,
2. an edge extractor,
3. a blur analysis to estimate distance.



1.3.1 - Preprocess

In this step, the picture data is reduced so that next steps of algorithm use less memory and thus less CPU time.

Technology to help blind people : Image restitution by sonorous signals

The color picture is translated into a gray level picture. Each pixel of the picture with red, green and blue data becomes a gray level thanks the following formula. This formula preserves human eyes feature where green is more detected than red and red is more detected than blue.

$$gray = (11*red + 16*green + 5*blue) / 32$$

One improvement for futur should be to introduce color variation between adjacent pixels. Indeed two adjacent pixels with different colors could be well discerned by human eyes on color picture and on the other hand, this two pixels could have same gray level after preprocess.

In this place, some picture process can be done to remove or attenuate video artefacts with algorithms. The artefact that seems to introduce the most amount of errors is light saturation.

1.3.2 - Edge detection

At this step, the algorithm must detect significant edges. An edge is an important transition between two areas. What means important ? It's a subjective point of view. Important could mean high transition between areas but it could also mean transition between two big areas.

Thus, the system has to detect hard transitions but not only that ones. It must detect smooth transitions between important areas also. It is very important because blurred transition is always smooth and these transitions MUST be detected. Transition between two adjacent pixels is not sufficient to detect all important edges.

A multiscale analysis has the ability to solve the problem. The picture is decayed into different scales. A transition between two pixels in higher scale¹ could be smooth while this transition at lower scale becomes hard. This can detect high contrast between two areas even if transition between them is smooth on original picture.

A threshold value is specified to extract edges. If this value is low, a lot of edges are detected but most part of them is not meaningful. On the other hand, if this value is high, few edges are detected but they are more meaningful.

1.3.3 - Distance estimation from blur level

To detect distances into a picture, the level of blur is computed. Indeed, for a picture focalized on near objects, as high is the blur level, as high is the distance from the lens.

To compute the blur level, the multiscale analysis from previous step is also used. Here is the process description in brief : when high contrast appears at a specific scale, it means that at this scale, the transition becomes hard and thus, the blur disappears. How high is the first scale (from higher to lower scale) where blur disappears, how high is the level of initial blur.

¹ A high scale offers more details than a low scale. For example, scale 1:10000 offers more details than scale 1:100000 on a map and 1:10000 (1 divided by 10,000) is higher than 1:100000.

Technology to help blind people : Image restitution by sonorous signals

Unfortunately, the blur detection is not very accurate. It can introduce some estimation mistakes. The trouble in blur detection can be distributed in three sets : features of captured objects, defaults or limitations of video hardware and problem from the algorithm itself.

Filmed objects problem.

Some features of objects in nature can introduce some « natural » blur. This blur is not produced by lens at all. Here are some examples :

- curve in subject (example : human face with nose and cheeks),
- smooth color change,
- high light that can introduce smooth luminosity variations on object's area,
- fog,
- air movements (heat).

These features introduces blur, then those objects are detected as far even they're near. Unfortunately, it seems to be very difficult to solve the problem.

Hardware limitation.

Video capture system will introduce some artefacts. Even if hardware is at top of technology, it always exists limits than can be crossed. Here are some encountered problems :

- blur from relative movement (camera movement or captured object movement),
- too low resolution,
- too low light, it erases edges,
- too much light (saturation), it introduces false hard transitions.

Parameters adjustment on camera (focus, aperture, capture time,...) can reduce trouble but it's highly probable that some problems will remain. We must find the most efficient choice between hardware possibility, hardware cost and error probability. Impact of these choices on the algorithm must be studied.

Problems from the algorithm.

The basis of the algorithm is multiscale analysis through a discrete wavelet transform (DWT). There's a lot of works on DWT, the algorithm uses some of them but with more simplification. This high level of simplicity is good for speed but bad for accurate. Some estimation mistakes can occur but the first tests with pictures in real conditions (parameter adjustment is poor) give usefull results. The most problem seems to be light saturation on over-exposed areas.

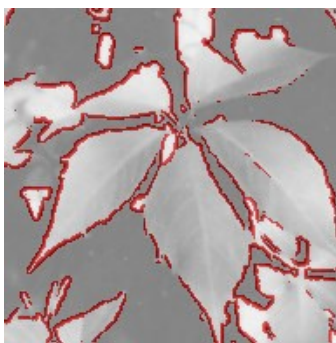
Another important problem comes from edge detection. One edge is detected in different scales but the detection at one scale can be slightly different than the detection of the same edge at another scale as you can see in the sample below. Red line is edge detection on higher scale (b) and blue line is edge detection on lower scale (c), red and blue line are not exactly at

**Technology to help blind people :
Image restitution by sonorous signals**

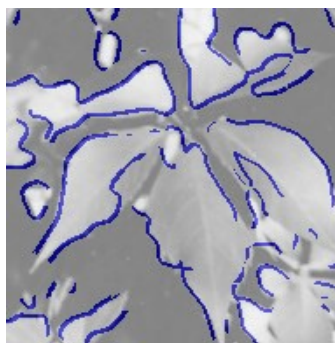
same positions (d). The detected edges at lower scale (c) are not always on the « real » edge and therefore can induce mistakes. The corresponding trouble is that those edges are detected as far because there is no sharp transition at this position. It seems not to be very important if some edges are nevertheless detected as near.



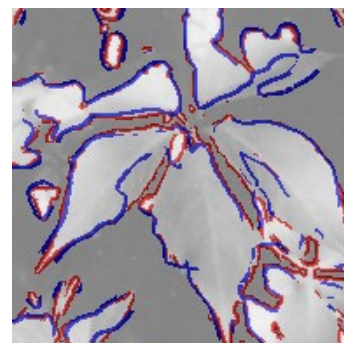
(a)



(b)



(c)



(d)

The algorithm tries to detect 4 zones (nearest, near, far, farest). Even if the detection is not very accurate, it can provide a good probability of distance : if a lot of pixels are computed as « nearest », there's a high probability that there's a near object. Note that the most important is to detect the nearest objects to avoid crash with them.

1.4 - Sonorous rendering

The result of process is edge detection into a picture and their distance. The number of pixels belonging to an edge is usually much more lesser than the total number of pixels. This relative few number of pixels can be arranged in 4 sets : nearest, near, far and farest points.

The blind people would like to « feel » these points. The first and most intuitive approach is to render points with tactile hardware. A vector of 2D sticks could represent the picture, sticks can be up or down, up sticks are pixels from edges. Unfortunately, this solution is very difficult and expensive to achieve. Another way is to use sonorous information to render pixels.

**Technology to help blind people :
Image restitution by sonorous signals**

A sound can be distinguished by some parameters, the variations of these parameters can induce a lot a different sounds. This large number of variations can be used to represent variations into picture.

Here are the characteristics of a sound :

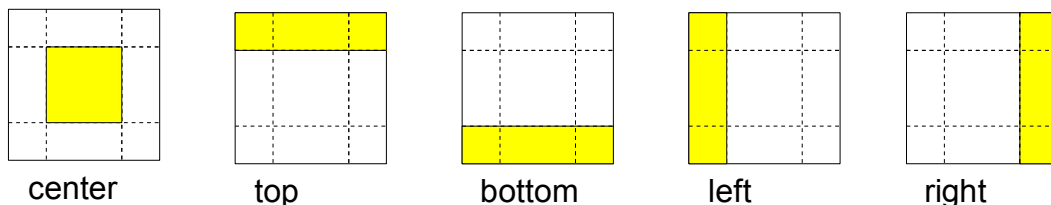
- frequency (note),
- amplitude (volume),
- tone (instrument),
- duration,
- position (stereo).

Note : we must take care about psychoacoustic limits of human ear.

1.4.1 - Rendering areas

Just one part of the picture can be rendered. A big area to render has the advantage to have a global idea about surrounding but the disadvantage is a poor accurate. If the user can chose dynamically the size and position of the picture's area to render, he has the choice between unaccurate global vision on large part or high precision on a small part.

The picture area to render can be divided into 5 subareas (see picture) : center, top, bottom, left and right subareas. This help the user to feel edge directions. Size of this subareas can be choosen by user.



Each subarea is rendered separatly. For each subarea, the number of nearest pixels, the number of near pixels, the number of far pixels and the number of farest pixels induce a particular sound.

1.4.2 - First approach

A first choice of sound to render picture is here fixed. It's not probably the best choice but it will be sufficient for first tests. Here are the parameters :

- **distances** : 4 frequencies (notes), one for each of the 4 distances,
- **point density** : the amplitude (volume) of note express the number of points for one distance,
- **subarea** : continuous sound for central subarea, sound with a duration of 200ms for surrounding subareas, the tone (instrument) distinguishes the subareas.

**Technology to help blind people :
Image restitution by sonorous signals**

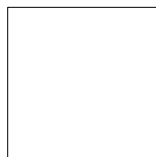
1.5 - Rendering parameters

As the user has a big job to interpret sound information, it is desirable to leave him the choice of most parameters. Some parameters are fixed most of the time, like precision for example and on the other hand, some ones change continually, like position of area to render for example.

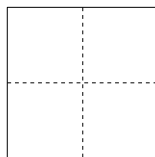
The user can select one of the suggested values for threshold in edge detection. Remember that if more edges are detected, most of them are not significant.

The user can select one of three subareas' distributions :

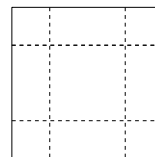
- only central subarea (a), no surrounding subareas,
- only surrounding subareas (b), no central subarea,
- central and surrounding subareas (c).



(a)



(b)



(c)

The user can select the size of the rendering area. The shape of the area is a square.

The user can decide the position of the area to render. The position is the center of the area all over the picture.

The user can chose to render one, more or all of the 4 distances.

Technology to help blind people : Image restitution by sonorous signals

2 - Project convenience

This short section present features of the project that brings something more than other projects. This evaluation takes into account on some disadvantages of the project. Note that blind user will have a conclusive opinion about usefulness of the project.

The input data is a picture like human eyes can see, it's an important feature that data to process by computer is the same as for human.

The main feature of the project in comparison with same kind of other projects is to provide a 3D information. In some projects, 3D data comes from sonar hardware or from a pair of cameras. These solutions are *a priori* more accurate than blur estimation but they are much more difficult to implement. The blur computation is less accurate but the vision is larger than with sonar hardware and is more simple than a couple of cameras.

2.1 - Advantages

Here some advantages in favour of the project :

- The data is basic with few computer interpretation.
- Large vision in 3 dimensions.
- No specific hardware, thus cheap and up-to-date hardware can be found (mobile, pocket PC,...). This hardware can be used for other uses.
- The system is able to replace sonar applications in presence of light.
- The picture process algorithm can be coupled with other video applications (OCR, color detection,...).

2.2 - Disadvantages

Here some disadvantages for the project :

- Picture process could be slow.
- Blur analysis is not very accurate and thus distance estimation either.
- Few possibilities in sonorous rendering and thus few accurate.
- There is currently no adequate material for the project (pending problem).
- The system doesn't work when there is too few light.

Technology to help blind people : Image restitution by sonorous signals

3 - Progress report

An application that processes picture runs for edge detection and blur computing. A lot of tests have to be done to validate the algorithm.

There is no specific hardware for input or output data. This step still needs research, there is no solution to invent but just material to be found.

3.1 - Video capture

This step depends highly of picture process hardware. If the hardware is a PC, a web cam could be fine. If it is a mobile phone (smartphone), the camera of the phone is the immediate solution.

This hardware must be adapted to provide a very short depth of field and a fixed focalisation on near objects, so that far objects are blurred on picture. This adaptation can be done with specific additional lens.

3.2 - Picture analysis

This point is succeeded the most. A first prototype, written in JAVA language, runs correctly. The JAVA language is able to run on a lot of operating systems.

Here the remaining points to study :

- hardware selection on which algorithm runs,
- connection with video capture hardware.

Computing time and **memory space** used are the most critical. The algorithm will perhaps be optimized for specific operating system.

Another important point is **autonomy**. The hardware must have light weight and low electric consumption. The system must run during several hours, ideally during the whole day.

The algorithm seems to run well, particularly for edge detection but a lot of tests has to be done yet to validate and improve solutions. Light saturation is still a problem. One futur improvement could be a better color detection to refine edge detection.

3.3 - Sonorous rendering

There is no hardware difficulty for sonorous rendering : a simple ear-phone is sufficient. To generate sounds, the application uses MIDI interface. This guarantees that it can run on a lot of different systems generating MIDI sounds.

On the other hand, there is no in-depth research yet about psychoacoustic limitation. Here again, a lot of tests is needed.

Technology to help blind people : Image restitution by sonorous signals

3.4 - Parameters control

A lot of parameters can be fixed by user. Unfortunately, at this level, only no blind people can use prototype because mouse is used to change parameters on prototype application.

Here parameters :

- 4 **distances** to render following 4 checkboxes,
- mouse position and left click put rendering **position** into picture,
- mouse wheel to change rendering area **size**,
- right click to change **subareas** type,
- a slider to select **precision**.

To use keyboard or specific hardware instead of mouse in the algorithm is not a very difficult work, it depends on the complexity of hardware driver. On the other hand, it's much more difficult to signal parameter value to blind people. Two kind of solutions are considered : a sonorous information or a specific hardware with tactile information.

Like edge rendering, parameter's information can be rendered with sound or words. This is the easiest solution to implement but more sounds could generate cacophony.

Specific hardware could provide tactile information. This is a more natural and more accurate solution. Here example of solution :

- **distances** : pushbuttons,
- **area position** : joystick or touchpad,
- **area size** : wheel (like for volume adjustment),
- **subareas type** : switch,
- **precision** : switch.

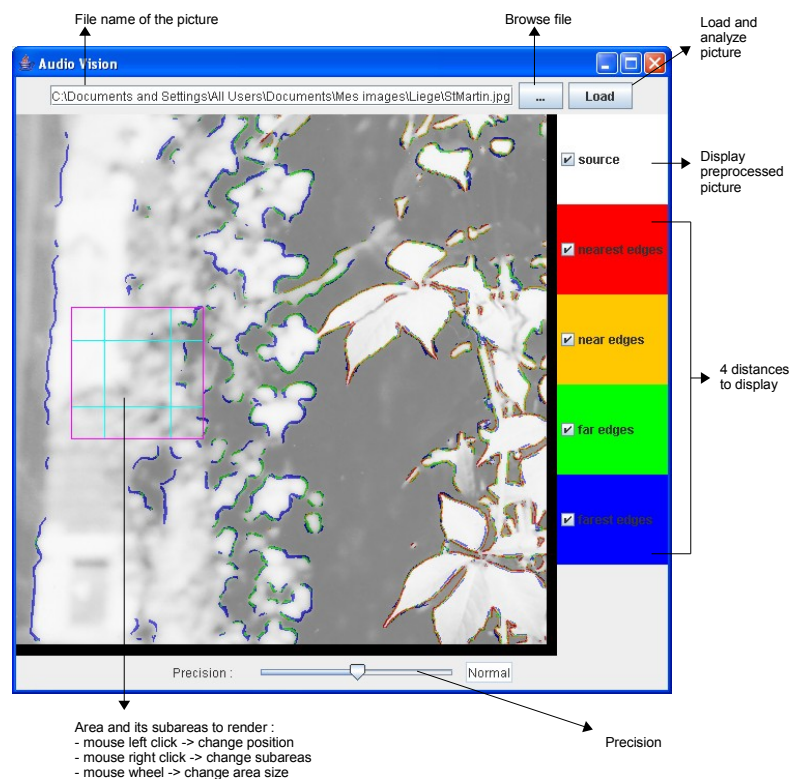
It is not certain that it is possible to find such hardware in market. Furthermore, a least one hand is needed to fix parameters.

The best solution is perhaps a mix of sonorous solutions and tactile solutions. Parameters that need to be changed often (like rendering position) should be tactile, parameters often fixed should be rendered by sounds.

Technology to help blind people : Image restitution by sonorous signals

4 - Prototype

A small application written in JAVA language is running with the algorithm explained above. A picture with out of focus areas can be loaded manually, edges are extracted and distances are computed. Then, sounds are issued with those informations. Unfortunately, the application cannot be used by blind people at the moment because most of parameters are fixed by mouse.



A setup with samples and documentation is available for Windows systems.

NOTE : Java virtual machine (JVM) must be installed to run the application.

5 - Conclusion

This brief paper presents quickly the project and its progress. The aims of this document is to receive some help to finalize the project, as advices or thorough co-operation. Advices from a specialist in the domain of technology for the blind people could help the project very much.

There is not yet financing for this project, it is too recent and too much little developped. Despite this fact, this paper tries to prove that the project is realizable and that it is usefull. If this document can convince of financials to finalize and market the product, it will have achieved its goal.

<p style="text-align: center;">Technology to help blind people : Image restitution by sonorous signals</p>

6 - Usefull links

Here are some Internet links to allow to compare project with current research in the field of technology for blind people.

The most interesting link is the following : <http://www.seeingwithsound.com> . The author created a system to render picture by sound. The system in this paper could be a complement of that one.

The site has a good explanation about sound rendering of picture :
<http://www.seeingwithsound.com/im2sound.htm>

Finally, the site offers a lot of links about in progress or commercial solutions to help blind people : <http://www.seeingwithsound.com/link.htm>

About picture process algorithm, the following links give information on basis of algorithm (wavelet transform) : first a simple but very friendly introduction to wavelet in computer graphics (<http://grail.cs.washington.edu/projects/wavelets>) and then a more general site about signal processing with wavelet (http://cas.ensmp.fr/~chaplais/Wavetour_presentation).