C-MIDI stands for MIDI that “sees” the notes. Video recognition of pianist hands and fingers

Gorodnichy, D.; Yogeswaran, A.; Comeau, G.

Publisher’s version / Version de l'éditeur:

4th Toronto-Montreal Computer Vision Workshop [Proceedings], 2006

National Research Council Canada
Conseil national de recherches Canada

Institute for Information Technology
Institut de technologie de l'information

NRC · CNRC

*C-MIDI Stands for MIDI That "Sees" The Notes **

Gorodnichy, D., Yogeswaran, A., Comeau, G.
May 2006

* published at the 4th Toronto-Montreal Computer Vision Workshop. Ottawa, Ontario, Canada. May 29-30, 2006. NRC 48786.

Canada

4th Toronto-Montreal Computer Vision Workshop, Ottawa, Canada, May 29-30, 2006.

C-MIDI stands for MIDI that "sees" the notes. Video recognition of pianist hands and fingers

Dmitry Gorodnichy[1], Arjun Yogeswaran[2], Gilles Comeau[3]
[1] Institute for Information Technology, National Research Council Canada
[2] School of Information Technology and Engineering, University of Ottawa
[3] Department of Music, University of Ottawa
http://synapse.vit.iit.nrc.ca/piano
http://www.piano.uottawa.ca

Summary:

Current music recording and transmitting technology allows teachers to teach piano remotely or off-line. This is in many cases the only way to teach music, especially in rural or distant areas where the ratio of piano teachers to piano students is extremely low [1]. MIDI recording technology allows a teacher to play a piano at one place and to see a piano played by itself, as by an ``invisible teacher", at another place (see Figure 1: the piano keys are pressed exactly at the same place, velocity and duration on a remote piano [2]. However, to know *how* these keys were played by a teacher remains unknown. This includes the knowledge of which hand played a key, which finger was used, and who (in case of a four hand musical piece) was playing. With the current advances in computer vision and video recognition, some of this knowledge can now be also transmitted [3].

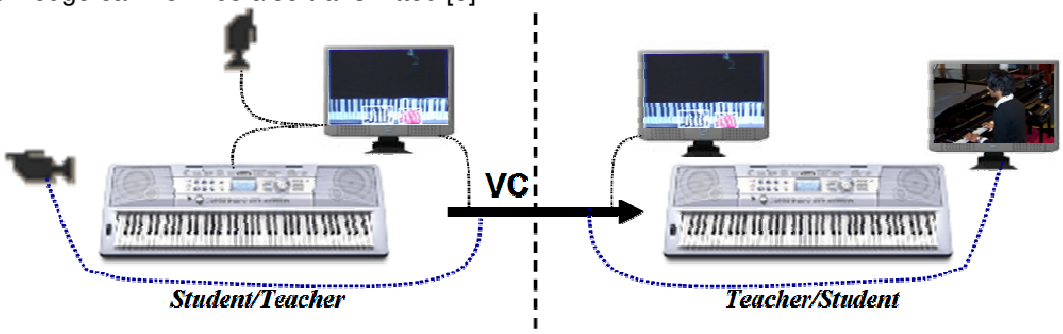


Figure 1. Video-conferencing (VC) for distant piano learning. A conventional session includes the transmission of a video image only (thick line). Video recognition technology allows one to transmit also the annotated video image (thin line).

This presentation summarizes a joint effort of the Computational Video Group of the Institute for Information Technology, National Research Council Canada and the Piano Pedagogy Lab of the Department of Music, University of Ottawa on the developing of a video recognition tool called ***C-MIDI*** that allows one to detect pianist hands and fingers using a video camera mounted on top of the piano keyboard (see Figure 2). By synchronizing video data with MIDI data, the program aims at annotating MIDI events according to the visual labels Hand={left,right} and Finger={1,2,3,4,5}. The result of the obtained video annotation of piano playing can then be shown on a computer screen for further perusal by a piano teacher or a student.

Besides distant and offline learning, the applications of presented technology for piano teaching are seen in automated video annotation, where it can be used for storing detailed information regarding music pieces in searchable databases (such as in [1]). It can also be used to facilitate producing music sheets and be used for score driven synthetic hand/finger motion generation (as in [4]). The obtained results show also the promise of performing vision-based annotations of MIDI playing for other than piano musical instruments. In particular, annotation of guitar and violin playing would be very useful, because

the recorded MIDI data of these instruments are very difficult to use for pedagogical needs, since they do not provide information on *how* and *where* on a fret- or finger- board these data are played.

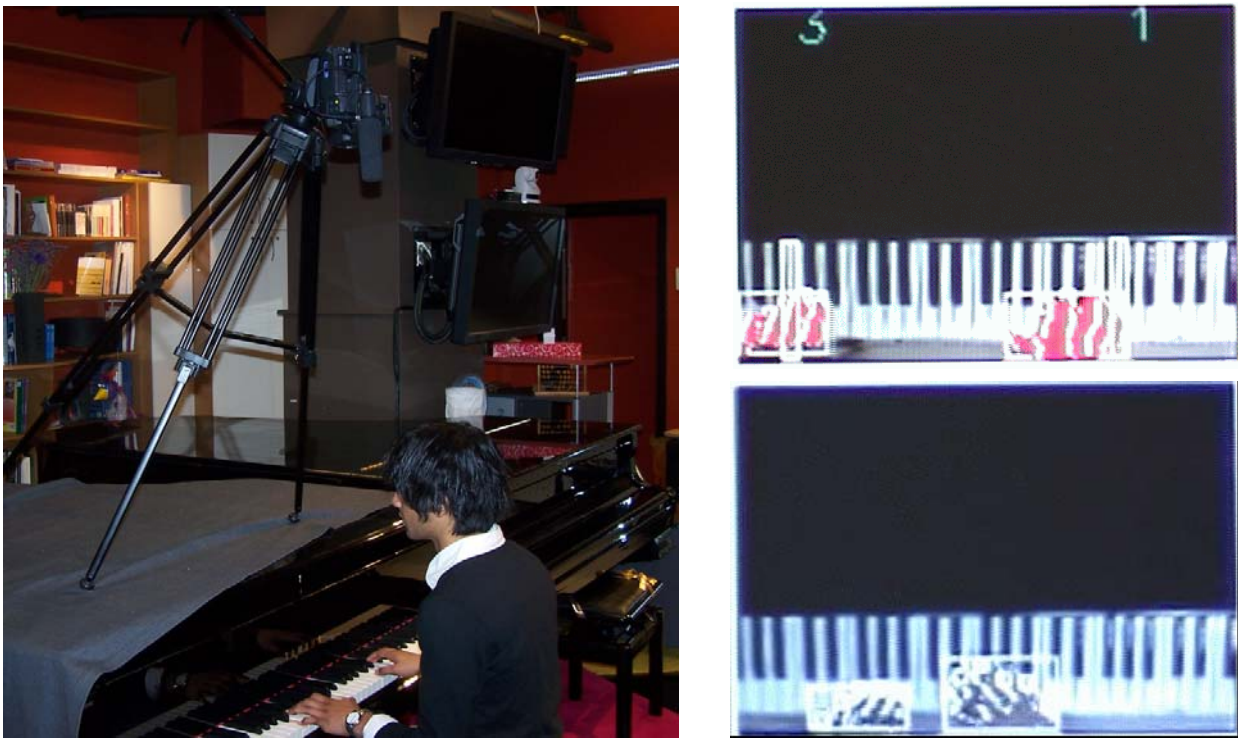


Figure 2: Setup for tracking pianist hands and fingers using a video camera in a professional piano studio environment with MIDI equipped grand piano and the video annotation of hands playing. Hands and finger edges are detected using video recognition techniques [3]. When a MIDI signal is received, meaning that a piano key was pressed, the hand and finger which are believed to press the piano key are shown: hand is highlighted in red, the finger number is shown on top of the image.

Acknowledgements:

This work is partially by SSHERC and CFI grants. MIDI events reader coding is done with the help of Mihir Sharma (SITE, UofO). Inspiring discussions with team members Emond Bruno and Martin Brooks (IIT, NRC) are also acknowledged.

References:

[1] B. Emond and M. Brooks. **The private video sharing and annotation server: A broadband application for teacher training and music education.** International Lisp Conference, New York, NY, 2003.

[2] International MIDI association, 1988. Standard MIDI Files 1.0. Los Angeles, CA.

[3] Dmitry O. Gorodnichy and Arjun Yogeswaran. "**Detection and tracking of pianist hands and fingers",** In Proc. of Canadian conference Computer & Robot Vision (CRV'06), Quebec city, Canada, June 7-9, 2006. Publisher: IEEE Computer Society Press - **http://www.computer.org/proceedings**. NRC 48492.

[4] H.Sekiguchi and S. Eiho. Generating the human piano performance in virtual space, 2005. International Conference on Pattern Recognition (ICPR'00), pp. 4477-4481.

C-MIDI stands for MIDI that "sees" the notes. Video recognition of pianist hands and fingers.

Dmitry O. Gorodnichy *, Arjun Yogeswaran**, Gilles Comeau***

* Institute for Information Technology, National Research Council Canada

** School of Information Technology and Engineering, University of Ottawa

*** Department of Music, University of Ottawa

http://synapse.vit.iit.nrc.ca/piano

4th Toronto-Montreal Computer Vision Workshop
May 28-29, 2006, Ottawa

Goals

1. To Recognize pianist hands (Left or Right) and fingers (1,2,3,4,5), as s/he plays a piano.
2. To "see" each (otherwise "blind") MIDI event:
   - old way: pitch, volume, etc
   - now way : + person, hand, finger

Examples of applications:

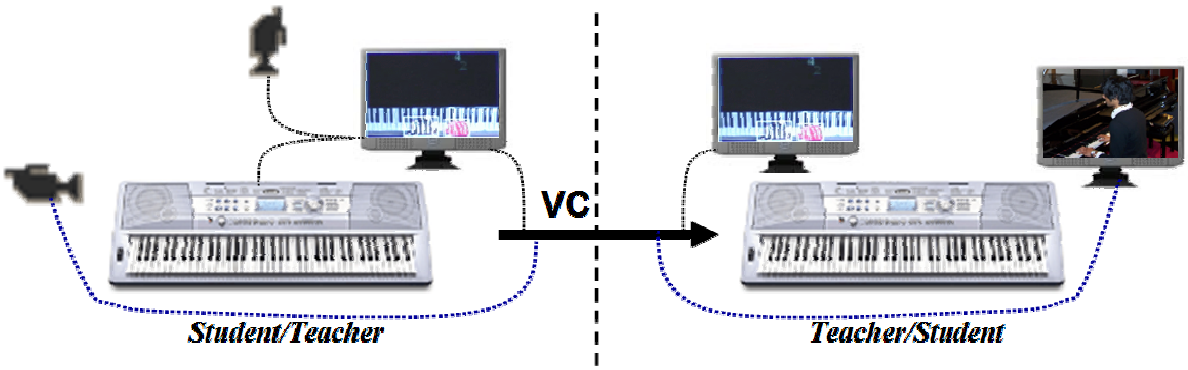
1. **Intelligent MIDI record /replay: "Play Midi of the left hand only"**
2. **Write finger number (suggestion) on top of each played note**
3. **Augmentation & Virtualization of piano performance**

Motivation: for Computer Vision

Unique unbiased testbed for hand/finger detection.

- In other applications (HCI, robotics, sign language), hands and fingers move *in order to be detected* (i.e. to send *visual information).*
- In piano playing, performer doesn't care about the visual information. Hands and fingers are extremely flexible & have unlimited set of states.

Motivation: for Music Teaching



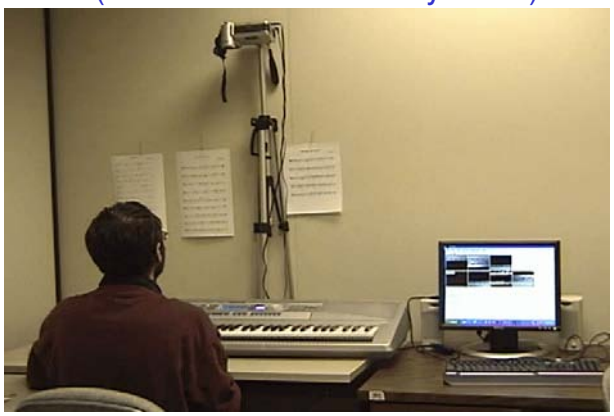
1. Video-conferencing (VC) for distant piano learning.
   - Conventional session includes transmission of a video image only.
   - Video recognition technology allows one to transmit also the annotated video image.

Also for:

2. for storing detailed information regarding music pieces
3. searchable databases (as in [4])
4. facilitating producing music sheets.
5. score driven synthetic hand/finger motion generation (as in [9])

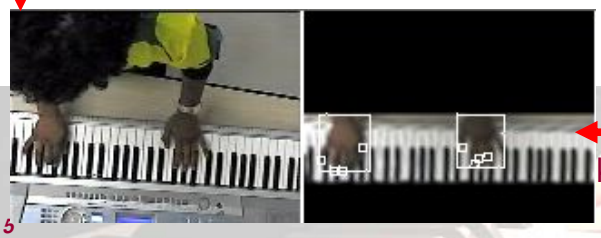
Setup, Video Input, Recognition Output

In home environment
(with Yamaha MIDI-keyboard)



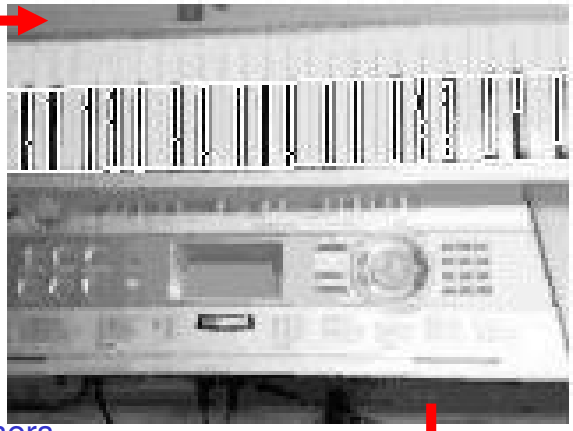
In Piano Pedagogy studio lab
(with MIDI-equipped grand piano)

Camera view from above

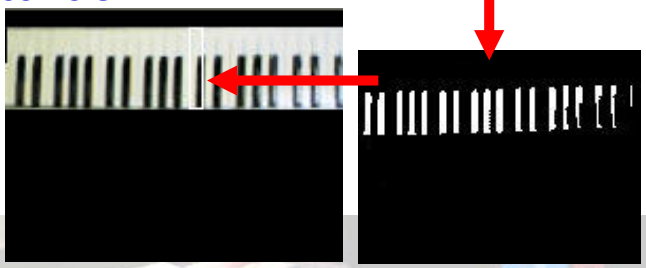


What computer would do:
keyboard rectification, key recognition
hand detection, finger detection

Step 1a. Image rectification

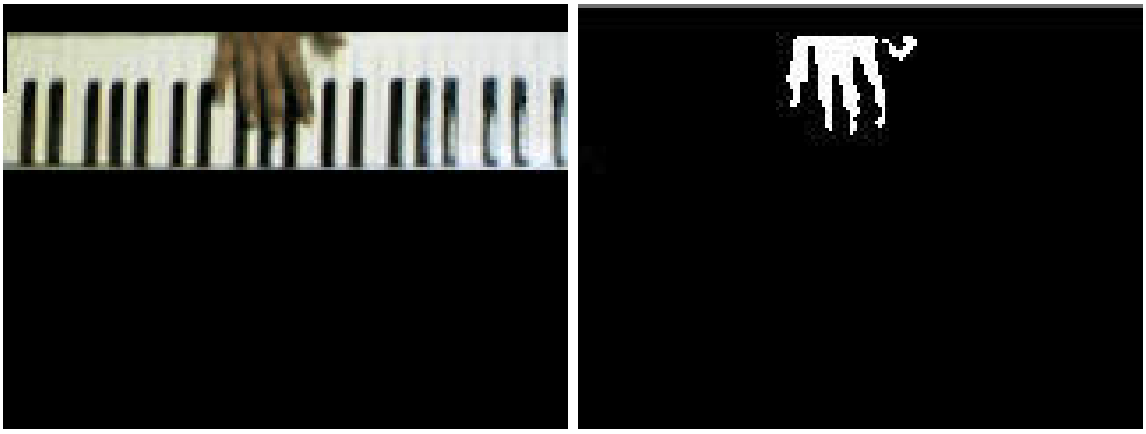


1. Top and bottom black corners are detected – lowest and highest points of black dilated blobs satisfying ratio.
2. Two lines are fit into detected corners.
3. Image rotated to make these lines parallel to Ox, cutting the image part.
4. Black blobs are counted to detect “C”.



Step 1b. Recognizing “C” key

Step 2a . Hand detection



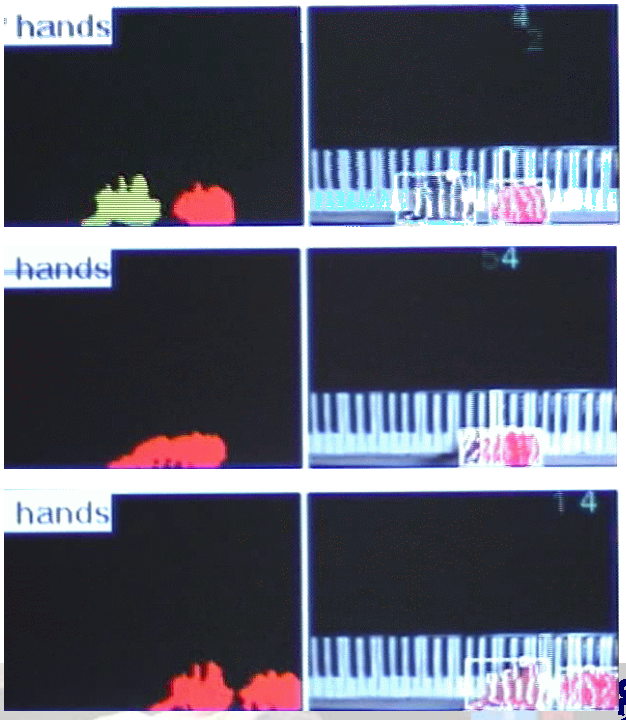
1. Background model of (detected in step 1) piano is maintained: IBG / DBG += new data AND I(no motion over several frames*)
2. Hands are detected as foreground: FG = |I − IBG| > 2 *DBG.
3. When they are detected,
   - Skin model is updated (UCS-masked HCr 2D histogram)
   - Number of hands is detected by K-means clustering

Step 2b. Hand Tracking

Technique:

Deformable box-shape template,

1. where only gradual changes (x,y,h,w, Vx,Vy) are allowed (compared to previous frame)
2. initialized by
   - foreground detection, or
   - skin colour tracking (by backprojecting 2D histogram of HCr learnt in Step 2a)



Foreground detection extracts blobs corresponding to hand images (left column)
Hand template tracking allows one to detect partially occluded hands (right column)

Step 3. Detecting fingers

Pianist fingers:

- *Unlike in other applications*, these fingers are never protruded! - Mostly bent towards keyboard (away from camera), often touching and occluding each other, tightly grouped together
- Low resolution video ➔ even more difficult to separate them

However: in camera these fingers are seen as convex objects!

Once hands are detected, fingers are detected by a new edge detection technique that scans hand areas searching for crevices.

Crevice-detection operator

NEW!

Conventional edge detection (Canny, Harris) don't use a-priory information about finger shapes => Return too many / small # pixels .

Definition: Crevices are locations in image where two convex shapes meet.

Finger edges are detected using crevice detection operator.

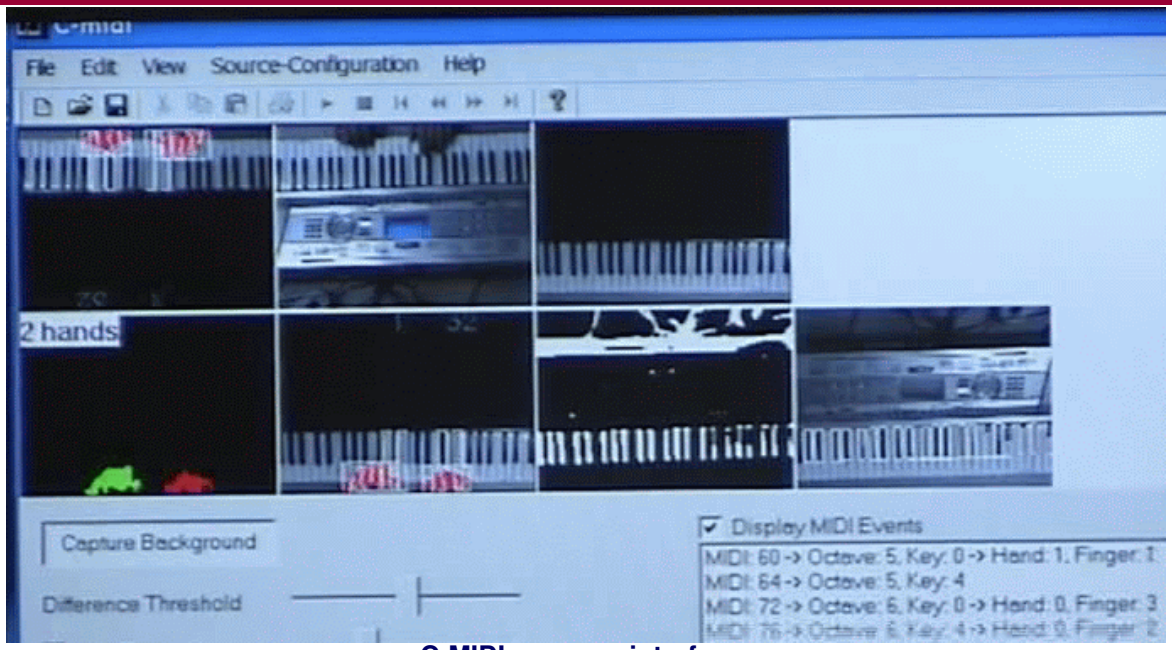
Crevice Detection operator:

scans in a one direction I(x) and marks a single pixel x*, where I(x) after going down goes up.

Requires post-processing:

- Method 1: merging adjacent pixels
- Method 2: filling a blob in between two "crevice edge pixels" x1* and x2* on the same line

Stage 4. Associating to MIDI events



C-MIDI program interface

When a MIDI signal is received (i.e. a piano key was pressed), the hand and finger that are believed to press the piano key are shown.
(hand is highlighted in red, the finger number is shown on top of the image).

C-MIDI output

Subwindows on GUI screen show (in clockwise order from top):

- image captured by camera;
- computed background image of the keyboard (used to detect hands as foreground);
- binarized image (used i. to detect black keys, ii. for video-MIDI calibration)
- automatically detected piano keys (highlighted as white rectangles),
- segmented blobs in foreground images (coloured by # of blobs)
- final finger and hand detection results (shown upside down, as camera views - on top left, and vertically flipped for viewing by a pianist - bottom middle)
  - The label of the finger that played a key is shown on the top of the image.
- results of vision-based MIDI annotation (in separate window at bottom right): each received MIDI event receives visual label
  - for hand (either 1 or 2, i.e. left or right)
  - for finger (either 1,2,3,4, or 5, counted from right to left) that played it. When the finger can not be determined, the annotation is omitted.

DEMO (Recorded LIVE)

- Three music pieces
  - of increasing complexity (speed, finger/hand motion)
  - played by professional piano teacher.

Limitations

- Temporal boundaries:
  - Video process practically real-time: annotating MM 160 1/8 notes (and faster) is possible
- Spatial boundaries:
  - 4 (5) octaves, small (10-year olds) hands – borderline
- Behavioral:
  - Overlapping of hands, overlapping by head, etc
- Environmental (Lighting, Shadows, Colours):
  - On different pianos, different auditoriums, different hands colour

Summary

- Piano keyboard detected, middle "C" recognized ...but can be improved.
- Hand segmentation required static setup ...but can be improved
- Hand Tracking is very good under certain conditions (no long overlapping, good light) ...but can be improved
- Finger Detection is most challenging, but the results are very encouraging ...and can be improved
- Integration with MIDI is done. But how to play the annotated MIDI back? ← Future work
- Original Publication. Theory and references:
  - Dmitry O. Gorodnichy and Arjun Yogeswaran. "**Detection and tracking of pianist hands and fingers",** In Proc. of Canadian conference Computer & Robot Vision (CRV'06), Quebec city, Canada, June 7-9, 2006. NRC 48492.
- Acknowledgements:
  - Partially supported by SSHERC and CFI grants.
  - MIDI events reader coding helped by Mihir Sharma (SITE, UofO)
  - Team member influences: Emond Bruno and Martin Brooks (IIT, NRC)